DOCUMENT RESUME

ED 471 518 TM 034 612

AUTHOR	Boston, Carol, Ed.
TITLE	Understanding Scoring Rubrics: A Guide for Teachers.
INSTITUTION	ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
ISBN	ISBN-1-886047-04-9
PUB DATE	2002-01-00
NOTE	97p.
CONTRACT	EDCO0032
AVAILABLE FROM	ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, 1129 Shriver Laboratory, College Park, MD 20742. Tel: 800-464-3742 (Toll Free); Web site: http://ericae.net.
PUB TYPE	Collected Works - General (020) -- Guides - Non-Classroom (055) -- ERIC Publications (071)
EDRS PRICE	EDRS Price MF01/PC04 Plus Postage.
DESCRIPTORS	*Classroom Techniques; Criteria; Elementary Secondary Education; *Feedback; Grades (Scholastic); Grading; *Performance Based Assessment; *Scoring Rubrics; *Student Evaluation

ABSTRACT

This compilation provides an introduction to using scoring rubrics in the classroom. When good rubrics are used well, teachers and students receive extensive feedback on the quality and quantity of student learning. When scoring rubrics are used in large-scale assessment, technical questions related to interrater reliability tend to dominate the literature. At the classroom level, concerns tend to center around how, when, and why to use scoring rubrics. The collection contains these chapters: (1) "Implementing Performance Assessment in the Classroom" (Amy Brualdi); (2) "An Introduction to Performance Assessment Scoring Rubrics" (Carole Perlman); (3) "Rubrics, Scoring Guides, and Performance Criteria" (Judith Arter); (4) "Scoring Rubric Development: Validity and Reliability" (Barbara M. Moskal and Jon A. Leydens); (5) "Converting Rubric Scores to Letter Grades" (Northwest Regional Educational Laboratory); (6) "Phantom of the Rubric or Scary Stories from the Classroom" (Joan James, Barb Deshler, Cleta Booth, and Jane Wade); (7) "Creating Rubrics through Negotiable Contracting" (Andi Stix); (8) "Designing Scoring Rubrics for Your Classroom" (Craig A. Mertler); and (9) "Scoring Rubrics Resources and Review" (Carol Boston). (Contains 11 tables, 9 figures, and 60 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 471 518

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

UNDERSTANDING SCORING RUBRICS

A Guide For Teachers

Edited by
Carol Boston

TM034612

BEST COPY AVAILABLE

Understanding Scoring Rubrics

A Guide For Teachers

Edited by

Carol Boston

Clearinghouse on Assessment and Evaluation
University of Maryland
College Park, MD

UNDERSTANDING SCORING RUBRICS:
A GUIDE FOR TEACHERS
Edited by
Carol Boston

ERIC Clearinghouse on Assessment and Evaluation
University of Maryland
1129 Shriver Laboratory
College Park, MD 20742
1.800.464.3742
http://ericae.net

Copyright ©2002
ERIC Clearinghouse on Assessment and Evaluation
All rights reserved.

Editor: Carol Boston
Design: Laura Chapman
Cover: Laura Chapman
Production: Printing Images, Inc.

This publication was prepared in part with funds from the Office of Educational Research and Improvement (OERI), U.S. Department of Education, and the National Library of Education (NLE) under contract ED99CO0032. The opinions expressed in this publication do not necessarily reflect the positions or policies of OERI, the U.S. Department of Education, or NLE.

Library of Congress Cataloging-in-Publication Data
Understanding scoring rubrics : a guide for teachers / edited by Carol Boston.
p. cm.
Includes bibliographical references.
ISBN 1-886047-04-9 (softcover : alk. paper)
1. Grading and marking (Students) I. Boston, Carol, 1961- II. ERIC Clearinghouse on Assessment and Evaluation.
LB3060.37 .U53 2002
371.27'2--dc21

2002002485

Printed in the United States of America

Contents

Introduction v

Chapter 1: Implementing Performance Assessment in the Classroom 1
Amy Brualdi

Chapter 2: An Introduction to Performance Assessment Scoring Rubrics 5
Carole Perlman

Chapter 3: Rubrics, Scoring Guides, and Performance Criteria 14
Judith Arter

Chapter 4: Scoring Rubric Development: Validity and Reliability 25
Barbara M. Moskal and Jon A. Leydens

Chapter 5: Converting Rubric Scores to Letter Grades 34
Northwest Regional Educational Laboratory

Chapter 6: Phantom of the Rubric or Scary Stories from the Classroom 41
Joan James, Barb Deshler, Cleta Booth, and Jane Wade

Chapter 7: Creating Rubrics Through Negotiable Contracting 66
Andi Stix

Chapter 8: Designing Scoring Rubrics for Your Classroom 72
Craig A. Mertler

Scoring Rubric Resources and Review 82
Carol Boston

References 86

INTRODUCTION

Scoring rubrics—guides that spell out the criteria for evaluating a task or performance and define levels of quality— are used in large-scale assessments, including the National Assessment of Educational Progress and some statewide assessments, as well as in many classroom tests and assignments. When scoring rubrics are used in large-scale assessments, technical questions related to inter-rater reliability (the likelihood that two or more raters assign the same score) tend to dominate the literature. At the classroom level, concerns tend to center around how, when, and why to use scoring rubrics.

Many teachers are interested in learning how to use scoring rubrics to assess student performance in the classroom. Scoring rubrics can help focus teachers and students on the most important elements of a learning task and encourage the metacognitive skill of self-assessment. They can also be helpful in reducing the subjectivity of some conventional grading methods.

To do classroom assessment well, teachers need to have an in-depth understanding of how knowledge is organized within their academic subjects, a clear picture of what students need to know and be able to do to demonstrate learning, and an awareness of how students typically progress in learning as they move from inexperienced or novice to experienced or master. *In Knowing What Students Know: The Science and Design of Educational Assessment*,[1] the National Academy of Sciences endorses student participation in assessment since "students learn more when they understand (and even participate in developing) the criteria by which their work will be evaluated, and when they engage

in peer and self-assessment during which they apply those criteria" (p. 7).

This compilation provides an introduction to using scoring rubrics in the classroom. The authors were selected on the basis of their sensitivity to the needs and realities of teaching and their ability to speak to both theory and practice. Because scoring rubrics go hand in hand with performance-based tasks, the opening chapter, "Implementing Performance Assessment in the Classroom" by Amy Brualdi, offers an overview of performance-based tasks and how they may be evaluated. Carole Perlman introduces performance assessment scoring rubrics in Chapter 2, including the difference between analytical and holistic rubrics and specific and general rubrics. She offers suggestions for adapting existing rubrics and creating original ones and introduces technical issues, particularly reliability in scoring.

In Chapter 3, "Rubrics, Scoring Guides, and Performance Criteria," Judith Arter points to the literature on how scoring rubrics can affect learning, particularly students' ability to assess their own work. Of particular interest in this piece is the metarubric, a means for evaluating the quality of a rubric. Chapter 4, "Scoring Rubric Development: Validity and Reliability" by Barbara M. Moskal and Jon A. Leydens, provides a discussion of content-, construct-, and criterion-related evidence for the validity of a scoring rubric and addresses issues of interrater and intrarater reliability.

When good rubrics are used well, teachers and students receive extensive feedback on the quality and quantity of student learning. The issue of then trying to feed this information into a more conventional grading system is complex and controversial. To illuminate the problem, we include in Chapter 5 sections of a module on converting rubric scores to letter grades from Northwest Regional Educational Laboratory's *Toolkit '98*. Chapters 6 and 7 describe classroom experiences of implementing scoring rubrics. The play "Phantom of the Rubric or Scary Stories from the Classroom" by Joan James, Barb Deshler, Cleta Booth, and Jane Wade follows preschool, elementary school, and middle school teachers through the highs and lows of a year of rubrics use. In "Creating Rubrics Through Negotiable Contracting," Andi Stix describes cases in which teachers have successfully involved their students in the creation of rubrics, thereby ensuring greater student engagement in subsequent products and performances. Chapter 8, "Designing Scoring Rubrics for Your Classroom," by Craig A. Mertler, helps readers apply their knowledge of scoring rubrics on a practical level. A final section of the book points to online resources related to scoring rubrics, including Web sites that offer extensive examples.

---

[1]Pellegrino, J. W., Chudowsky, N., and Glaser, R., eds. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy of Sciences. [Available online: http://www.nap.edu/books/0309072727/html/].

Chapter 1

Implementing Performance Assessment in the Classroom

Amy Brualdi
Lexington (MA) Public Schools

This chapter will introduce you to performance assessments, the type of student evaluation commonly associated with scoring rubrics. Performance assessments require that students demonstrate their knowledge and skills in context rather than simply complete a worksheet or a multiple-choice test. A performance assessment in science might require that a student conduct an experiment; one in composition might require that students write a letter to the editor about a real-world issue. Some performance assessments require that students apply multiple skills across subjects. Good questions to ask yourself before developing any assessment are, "What is the purpose of this assessment?" and "What do I expect my students to know and be able to do?" Once you've begun to specify the criteria for successful products or performances, you're on your way to creating a scoring rubric.

If you are like most teachers, it probably is a common practice for you to devise some sort of test to determine whether a previously taught concept has been learned before introducing something new to your students. Probably, this will be either a completion or a multiple-choice test. However, it is difficult to write completion or multiple-choice tests that go beyond the recall level. For example, the results of an English test may indicate that a student knows each story has a beginning, a middle, and an end. However, these results do not guarantee that a student will write a story with a clear beginning, middle, and end. Because of this, educa-

---

This chapter first appeared in the online, peer-reviewed journal, *Practical Assessment, Research & Evaluation*, 6 (2), available at http://ericae.net/pare/. It was written when Amy Brualdi was a staff member of the ERIC Clearinghouse on Assessment and Evaluation.

BEST COPY AVAILABLE

tors have advocated the use of performance-based assessments.

Performance-based assessments "represent a set of strategies for the . . . application of knowledge, skills, and work habits through the performance of tasks that are meaningful and engaging to students" (Hibbard and others, 1996, p. 5). This type of assessment provides teachers with information about how a child understands and applies knowledge. Also, teachers can integrate performance-based assessments into the instructional process to provide additional learning experiences for students.

The benefits of performance-based assessments are well documented. However, some teachers are hesitant to implement them in their classrooms. Commonly, this is because these teachers feel they don't know enough about how to assess a student's performance fairly (Airasian,1991). Another reason for reluctance in using performance-based assessments may be previous experiences with them when the execution was unsuccessful or the results were inconclusive (Stiggins, 1994). This chapter outlines the basic steps that you can take to plan and implement effective performance-based assessments.

Defining the Purpose of the Performance-Based Assessment

In order to administer any good assessment, you must have a clearly defined purpose. Thus, you must ask yourself several important questions:

- ✓ What concept, skill, or knowledge am I trying to assess?
- ✓ What should my students know?
- ✓ At what level should my students be performing?
- ✓ What type of knowledge is being assessed: reasoning, memory, or process? (Stiggins, 1994)

By answering these questions, you can decide what type of activity best suits you assessment needs.

Choosing the Activity

After you define the purpose of the assessment, you can make decisions concerning the activity. There are some things that you must take into account before you choose the activity: time constraints, availability of resources in the classroom, and how much data is necessary in order to make an informed decision about the quality of a student's performance (This consideration is frequently referred to as sampling.).

The literature distinguishes between two types of performance-based assessment activities that you can implement in your classroom: informal and formal (Airasian, 1991; Popham, 1995; Stiggins, 1994). When a student is being informally assessed, the student does not know that the assessment is taking place. As a teacher, you probably use informal performance assessments all the time. One example of something that you may assess in this manner is how children interact with other children (Stiggins, 1994). You also may use informal assessment to assess a student's typical behavior or work habits.

A student who is being formally assessed knows that you are evaluating him

or her. When a student's performance is formally assessed, you may either have the student perform a task or complete a project. You can either observe the student as he or she performs specific tasks or evaluate the quality of finished products.

You must beware that not all hands-on activities can be used as performance-based assessments (Wiggins, 1993b). Performance-based assessments require individuals to apply their knowledge and skills in context, not merely complete a task on cue.

Defining the Criteria

After you have determined the activity as well as what tasks will be included in the activity, you need to define the elements of the project or task that you will evaluate to determine the success of the student's performance. Sometimes, you may be able to find these criteria in local and state curriculums or other published documents (Airasian, 1991). Although these resources may prove to be very useful to you, please note that some lists of criteria may include too many skills or concepts or may not fit your needs exactly. With this in mind, be sure to review criteria lists before applying any of them to your performance-based assessment. When you develop your own criteria, Airasian (1991, p. 244) suggests that you complete the following steps:

- ✓ Identify the overall performance or task to be assessed and perform it yourself or imagine yourself performing it.
- ✓ List the important aspects of the performance or product.
- ✓ Try to limit the number of performance criteria so they can all be observed during a pupil's performance.
- ✓ If possible, have groups of teachers think through the important behaviors included in a task.
- ✓ Express the performance criteria in terms of observable pupil behaviors or product characteristics.
- ✓ Don't use ambiguous words that cloud the meaning of the performance criteria.
- ✓ Arrange the performance criteria in the order in which they are likely to be observed.

You may even wish to allow your students to participate in this process. You can do this by asking the students to name the elements of the project/task that they would use to determine how successfully it was completed (Stix, 1997).

Having clearly defined criteria will make it easier for you to remain objective during the assessment because you will know exactly which skills and/or concepts you are supposed to be assessing. If your students were not already involved in the process of determining the criteria, you will usually want to share them with your students. This will help students know exactly what is expected of them.

Creating Performance Rubrics

As opposed to most traditional forms of testing, performance-based assessments don't have clear-cut right or wrong answers. Rather, there are degrees to

which a person is successful or unsuccessful. Thus, you need to evaluate the performance in a way that will allow you take those varying degrees into consideration. This can be accomplished by creating rubrics.

A rubric is a rating system by which teachers can determine at what level of proficiency a student is able to perform a task or display knowledge of a concept. With rubrics, you can define the different levels of proficiency for each criterion. As with the process of developing criteria, you can either utilize previously developed rubrics or create your own. When using any type of rubric, you need to be certain that the rubrics are fair and simple. Also, the performance at each level must be clearly defined and accurately reflect its corresponding criterion (or subcategory) (Airasian, 1991; Popham, 1995; Stiggins, 1994).

When deciding how to communicate the varying levels of proficiency, you may wish to use impartial words instead of numerical or letter grades. Words such as "novice," "apprentice," "proficient," and "excellent" are frequently used.

As with criteria development, allowing your students to assist in the creation of rubrics may be a good learning experience for them. You can engage students in this process by showing them examples of the same task performed/project completed at different levels and discuss to what degree the different elements of the criteria were displayed. If your students do not help to create the different rubrics, you will probably want to share those rubrics with your students before they complete the task or project.

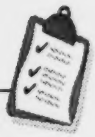
Assessing the Performance

You can give feedback on a student's performance in the form of either a narrative report or a grade. There are several different ways to record the results of performance-based assessments (Airasian,1991; Stiggins,1994):

- ✓ *Checklist approach.* When teachers use this, they only have to indicate whether or not certain elements are present in the performances.
- ✓ *Narrative/anecdotal approach.* When teachers use this, they write narrative reports of what was done during each of the performances. From these reports, teachers can determine how well their students met their standards.
- ✓ *Rating scale approach.* When teachers use this, they indicate the degree to which the standards were met. Usually, teachers will use a numerical scale. For instance, a teacher may rate each criterion on a scale of one to five, with one meaning "skill barely present" and five meaning "skill extremely well-executed."
- ✓ *Memory approach.* When teachers use this, they observe the students performing the tasks without taking any notes. They use the information from their memory to determine whether or not the students were successful. (Please note that this approach is not recommended.)

While it is a standard procedure for teachers to assess students' performances, teachers may wish to allow students to assess them themselves. Permitting students to do this provides them with the opportunity to reflect upon the quality of their work and learn from their successes and failures.

Chapter 2



An Introduction to Performance Assessment Scoring Rubrics

Carole Perlman
Office of Accountability, Chicago Public Schools

Chapter 1 provided an overview of performance assessment, including choosing the activity and defining the criteria to be used as the basis for assessing the performance. The criteria become the basis for a scoring rubric, checklist, narrative report, or rating scale. This chapter spells out questions to ask when you are developing performance assessment tasks so you can gather good information about student learning and challenge and engage your students in the process. You'll learn the difference between an analytical and a holistic rubric and receive step-by-step practical guidance about how to identify good existing rubrics, adapt them for your purposes, or create new ones, if necessary. The table will provide you with ideas about what features are commonly included in scoring rubrics for various subjects. The technical issues of validity and reliability are also raised in this chapter and dealt with in greater detail in Chapter 4.

Unlike a multiple-choice or true-false test in which a student is asked to choose one of the responses provided, a performance assessment requires a student to perform a task or generate his or her own response. For example, a performance assessment in writing would require a student to actually write something, rather than simply answer some multiple-choice questions about grammar or punctuation. Performance assessments are well suited for measuring complex learning outcomes such as critical thinking, communication, and problem-solving skills.

A performance assessment consists of two parts: a task and a set of scoring cri-

---

This chapter was derived from *The CPS Performance Assessment Idea Book*, written by Dr. Carole Perlman. ©1994 Chicago Public Schools. Used with permission. All rights reserved.

BEST COPY AVAILABLE

teria or "rubric." The task may be a product, performance, or extended written response to a question that requires the student to apply critical thinking skills. Some examples of performance assessment tasks include written compositions, speeches, works of art, science fair projects, research projects, musical performances, open-ended math problems, and analyses and interpretations of stories students have read.

Because a performance assessment does not have an answer key in the sense that a multiple-choice test does, scoring a performance assessment necessarily involves making some subjective judgments about the quality of a student's work. A good set of scoring guidelines or "rubric" provides a way to make those judgments fair and sound. It does so by setting forth a uniform set of precisely defined criteria or guidelines that will be used to judge student work.

The rubric should organize and clarify the scoring criteria well enough so that two teachers who apply the rubric to a student's work will generally arrive at the same score. The degree of agreement between the scores assigned by two independent scorers is a measure of the reliability of an assessment. This type of consistency is needed for a performance assessment to yield good data that can be meaningfully combined across classrooms and used to develop school improvement plans.

Selecting Tasks for Performance Assessments

The best performance assessment tasks are interesting, worthwhile activities that relate to your instructional outcomes and allow your students to demonstrate what they know and can do. As you decide what tasks to use, consider the following criteria that are adapted from Herman, Aschbacher, and Winters (1992):

Does the task truly match the outcome(s) you're trying to measure?

This is a must. It follows from this that the task shouldn't require knowledge and skills that are irrelevant to the outcome. For example, if you are trying to measure speaking skills, asking the students to orally summarize a difficult science article penalizes those students who are poor readers or who lack the science background to understand the article. In that case, you would not know whether you were measuring speaking or (in this case) extraneous reading and science skills. Sometimes it is possible to provide pertinent background material that would enable students to perform well on the task, despite deficiencies in prior knowledge. Allowing students access to textbooks and reference materials they know how to use may also be helpful.

Does the task require the students to use critical thinking skills?

Must the student analyze, draw inferences or conclusions, critically evaluate, synthesize, create, or compare? Or is recall all that is being assessed? The solution to the task should generally not be one in which the students have received specific instruction, since what is measured in that case may simply be rote memory. For example, suppose an instructional outcome included analyzing an author's point of view. If a class discussion is devoted to an analysis of the authors' points of view in two editorials and the students are then asked to write a composition analyzing the authors' positions

expressed in the same editorials, what is really being measured is probably recall of the class discussion, rather than the student's ability to do the analysis. A better assessment would be to ask the students to analyze some editorials that haven't been discussed in class.

Is the task a worthwhile use of instructional time?

Performance assessments may be time consuming, so it stands to reason that that time should be well spent. Instead of being an "add-on" to regular instruction, the assessment should be part of it.

Does the assessment use engaging tasks from the "real world"?

The task should capture the students' interest well enough to ensure that they are willing to try their best. Does the task represent something important that students will need to do in school and in the future? Many students are more motivated when they see that a task has some meaning or connection to life outside the classroom.

Can the task be used to measure several outcomes at once?

If so, the assessment process can be more efficient, by requiring fewer assessments overall.

Are the tasks fair and free from bias?

Is the task an equally good measure for students of different genders, cultures, and socioeconomic groups represented in your school population? Will all students have equivalent resources—at home or at school—with which to complete the task? Have all students received equal opportunity to learn what is being measured?

Will the task be credible?

Will your colleagues, students, and parents view the task as being a meaningful, challenging, and appropriate measure?

Is the task feasible?

Can students reasonably be expected to complete the task? Will you and your students have enough time, space, materials, and other resources? Does the task require knowledge and skills that you will be able to teach?

Is the task clearly defined?

Are instructions for teachers and students clear? Does the student know exactly what is expected?

Understanding Scoring Rubrics

A scoring rubric has several components, each of which contributes to its usefulness. These components include the following:

- ✓ One or more dimensions on which performance is rated,
- ✓ Definitions and examples that illustrate the attribute(s) being measured, and
- ✓ A rating scale for each dimension.

Ideally, there should also be examples of student work that fall at each level of the rating scale.

Analytical vs. Holistic Rating

A rubric with two or more separate scales—for example, a science lab rubric broken down into sections related to hypothesis, procedures, results, and conclusion— is called an analytical rubric. A scoring rubric that uses only a single scale yields a global or holistic rating. The overall quality of a student response might, for example, be judged excellent, proficient, marginal, or unsatisfactory. Holistic scoring is often more efficient, but analytical scoring systems generally provide more detailed information that may be useful in planning and improving instruction and communicating with students.

Whether the rubric chosen is analytical or holistic, each point on the scale should be clearly labeled and defined. There is no single best number of scale points, although it is best to avoid scales with more than six or seven points. With very long scales, it is often difficult to adequately differentiate between adjacent scale points (e.g., on a 100-point scale, it would be hard to explain why you assigned a score of 81 rather than 80 or 82). It is also harder to get different scorers to agree on ratings when very long scales are used. Extremely short scales, on the other hand, make it difficult to identify small differences between students. Short scales may be adequate if you simply want to divide students into two or three groups, based on whether they have attained or exceeded the standard for an outcome.

The rule of thumb is to have as many scale points as can be well defined and that adequately cover the range from very poor to excellent performance. If you decide to use an analytic rubric, you may wish to add or average the scores from each of the scales to get a total score. If you feel that some scales are more important than others (and assuming that the scales are of equal length), you may give them more weight by multiplying those scores by a number greater than one. For example, if you felt that one scale was twice as important as all the others, you would multiply the score on that scale by two before you added up the scale scores to get a total score.

Specific vs. General Rubrics

Scoring rubrics may be specific to a particular assignment or they may be general enough to apply to many different assignments. Usually the more general rubrics prove to be most useful, since they eliminate the need for constant adaptation to particular assignments and because they provide an enduring vision of high-quality work that can guide both students and teachers.

A rubric can be a powerful communications tool. When it is shared among teachers, students and parents, the rubric communicates in concrete and observable terms what the school values most. It provides a means for you and your colleagues to clarify your vision of excellence and convey that vision to your students. It can also provide a rationale for assigning grades to subjectively scored assessments. Sharing the rubric with students is vital—and only fair—if we expect them to do their best possible work. An additional benefit of sharing the rubric is that it empowers students to critically evaluate their own work.

In order for a rubric to effectively communicate what we expect of our students, it is necessary that students and parents be able to understand it. This may require restating all or part of the rubric to eliminate educational jargon or to explain a rubric in a way that is appropriate for the student's developmental level (e.g., "This story has a beginning, middle, and end" is clearer and more helpful than "Observes story structure conventions").

Selecting Scoring Rubrics

Teachers interested in using rubrics to assess performance-based tasks have three options: using an existing rubric as is, adapting or combining rubrics to suit the task, or creating a rubric from scratch. The resource section at the back of this book contains Web addresses of several sites offering sample rubrics.

If you're looking at existing rubrics, ask yourself these questions:

- ✓ Does the rubric relate to the outcome(s) being measured? Does it address anything extraneous?
- ✓ Does the rubric cover important dimensions of student performance?
- ✓ Do the criteria reflect current conceptions of "excellence" in the field?
- ✓ Are the categories or scales well defined?
- ✓ Is there a clear basis for assigning scores at each scale point?
- ✓ Can the rubric be applied consistently by different scorers?
- ✓ Can the rubric be understood by students and parents?
- ✓ Is the rubric developmentally appropriate?
- ✓ Can the rubric be applied to a variety of tasks?
- ✓ Is the rubric fair and free from bias?
- ✓ Is the rubric useful, feasible, manageable, and practical?

To adapt an existing rubric to better suit your task and objectives, you could:

- ✓ Re-word parts of the rubric.
- ✓ Drop or change one or more scales of an analytical rubric.
- ✓ Omit criteria that are not relevant to the outcome you are measuring.
- ✓ "Mix and match" scales from different rubrics.
- ✓ Change the rubric for use at a different grade.
- ✓ Add a "no response" category at the bottom of the scale.
- ✓ Divide a holistic rubric into several scales.

If adopting or adapting an existing rubric doesn't work for you, here are some steps to follow to develop your own scoring rubric:

1. With your colleagues, make a preliminary decision about the dimensions of the performance or product to be assessed. The dimensions you choose may be guided by national curriculum frameworks, publications of professional organizations, sample scoring rubrics (if available), or experts in the subject area in which you are working. (See Table 1 for a list of dimensions often used in scoring rubrics.) Alternatively, you and your colleagues may brainstorm a list of as many key attributes of the product/performance to be rated as you can. What do you look for when you grade assignments of this nature? When you teach, which elements of this product/performance do you emphasize?
2. Look at some actual examples of student work to see if you have omitted any important dimensions. Try sorting examples of actual student work into three piles: the very best, the poorest, and those in between. With your colleagues try to articulate what makes the good assignments good.
3. Refine and consolidate your list of dimensions as needed. Try to cluster your tentative list of dimensions into just a few categories or scales.

Alternatively, you may wish to develop a single, holistic scale. There is no "right" number of dimensions, but there should be no more than you can reasonable expect to rate. The dimensions you use should be related to the learning outcome(s) you are assessing.

4. Write a definition of each of the dimensions. You may use your brainstormed list to describe exactly what each dimension encompasses.
5. Develop a continuum (scale) for describing the range of products/performances on each of the dimensions. Using actual examples of student work to guide you will make this process much easier. For each of your dimensions, what characterizes the best possible performance of the task? This description will serve as the anchor for each of the dimensions by defining the highest score point on your rating scale. Next describe in words the worst possible product/performance. This will serve as a description of the lowest point on your rating scale. Then describe characteristics of products/performances that fall at the intermediate point of the rating scale for each dimension. Often these points will include some major or minor flaws that preclude a higher rating.
6. Alternatively, instead of a set of rating scales, you may choose to develop a holistic scale or a checklist on which you will record the presence or absence of the attributes of a quality product/performance.
7. Evaluate your rubric using the criteria discussed above.
8. Pilot test your rubric or checklist on actual samples of student work to see whether the rubric is practical to use and whether you and your colleagues can generally agree on what scores you would assign to a given piece of work.
9. Revise the rubric and try it out again. It's unusual to get everything right the first time. Did the scale have too many or two few points? Could the definitions of the score points be made more explicit?
10. Share the rubric with your students and their parents. Training students to use the rubric to score their own work can be a powerful instructional tool. Sharing the rubric with parents will help them understand what you expect from their children.

Table 1: Features of Student Work Commonly Included in Scoring Rubrics

Reading
- Summarize
- Integrate
- Synthesize ideas within and between texts
- Use knowledge of text structure and genre to construct meaning:
  - main ideas
  - summaries
  - themes
  - interpretations
  - literacy devices
  - multiple perspectives
- Use reading strategies
- Apply and transfer to new situations, problems, text
- Contributing skills:
  - decoding
  - structural
  - analysis
  - vocabulary
  - study skills

Writing
- Purposes
  - persuade
  - inform
- Features
  - integration
  - focus
  - support/elaboration
  - organization
  - conventions
- Processes
  - planning
  - drafting
  - revision
  - editing

Mathematics
- Problem-solving strategies
  - identify problems
  - apply strategies
  - use concepts, procedures, tools
- Representation
  - charts
  - graphs
- Reasoning
  - interpret
  - generalize
- Communication
  - clear
  - organized
  - complete
  - detailed
  - mathematical language, terminology, symbols, notations
- Content
  - number concepts and skills
  - percent/ratio proportion
  - measurement
  - algebraic concepts and skills
  - geometric concepts and skills

Science
- Investigations
  - hypotheses
  - other data
  - observe
  - use equipment
  - draw inferences
- Concepts/basic vocabulary of biological, physical, and environmental sciences
- Applications
- Social, environmental implications and limitations
- Communication
  - language
  - problem/issue
  - observation
  - evidence
  - conclusion/interpretations

Social Science
- Facts and concepts
- Critical thinking
  - issues
  - information
  - conclusions
- Significant personalities, terms, events
- Relationships within and across disciplines
- Communication
  - position
  - alternative interpretations
  - consequences
  - support
  - organization
  - conclusions
  - alternatives
- Group collaboration
  - participation
  - shared responsibility
  - responsiveness
  - forethought
  - preparation

Art
- Formal elements
  - structure
  - composition
- Technical
  - techniques
  - materials
- Sensory elements
- Expressive
  - mood
  - emotional energy/quality
- Identify elements
- Integration of elements
- Impact of elements

Physical Development/Health
- Human physical development and function
- Health principles
  - nutrition
  - stress management
  - exercise
  - self-concept
  - drug use and abuse
  - illness, prevention and treatment
- Apply to self
  - as consumer
  - as participant in sport and leisure activity
  - as life saver

(Quellmalz, 1991)

BEST COPY AVAILABLE

Using Scoring Rubrics: Some Technical Considerations

An assessment is valid for a particular purpose if it *in fact measures what it* was intended to measure. An assessment of a learning outcome is valid to the extent that scores truly measure that outcome and are not affected by anything irrelevant to the outcome.

Some important aspects of validity are content coverage, generalizability, and fairness. The assessments for a given outcome should be aligned with both the outcome and the instruction and, when taken together, should cover all important aspects of the outcome. The assessments should address the higher-order thinking skills specified in the outcome. The tasks used should have answers or solutions that can't be memorized, but which, instead, call on the student to apply knowledge and skills to a new situation.

Assessment results are generalizable to the extent that available evidence shows that scores on one assessment can predict how well students perform on another assessment of the same outcome.

An assessment is reliable if it yields results that are accurate and stable. In order for a performance assessment to be reliable, it should be administered and scored in a consistent way for all the students who take the assessment. Once you decide on a rubric, the best way to promote reliable scoring is to have well-trained scorers who thoroughly understand the rubric and who periodically score the same samples of student work to ensure that they are maintaining uniform scoring.

Another way to increase reliability is to try hard to stick to the rubric as you score student work. Not only will this increase reliability and validity, but it's only fair that the agreed-upon rubric that's shared with students and parents should be what is actually used to rate student work. Nonetheless, human beings making subjective judgments may unintentionally rate students based on things that aren't in the rubric at all. The conscientious scorer will frequently monitor his or her thinking to prevent extraneous factors from creeping into the assessment process. Table 2 on page 14 contains a list of some extraneous factors to watch out for.

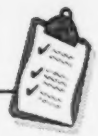
A good scoring rubric will help you and other raters be accurate, unbiased, and consistent in scoring and document the procedures used in making important judgments about students.

Table 2: Problems and Pitfalls Encountered by Scorers

Positive-Negative Leniency Error:	The scorer tends to be too hard or too easy on everyone.
Trait Error:	The scorer tends to be too hard or too easy on a given trait, criterion, or scale.
Appearance:	The scorer thinks more about how the paper or project looks than about the quality.
Length:	Is longer better? Not necessarily.
Fatigue:	Everybody gets tired.
Repetition Factor:	This paper is just like the last 50.
Order Effects:	If you've just read 10 bad papers, an average one may start to look like Shakespeare by comparison.
Personality Clash:	It's tougher if you don't like the topic or the student's point of view.
Skimming:	Doesn't the first paragraph pretty well tell the story? (Hint: No.)
Error of Central Tendency:	Using an odd-numbered scoring scale? Beware the dreaded "mid-point dumping ground."
Self-Scoring:	Are you a perceptive reader? Be sure what you're scoring is the writer's work—not your own skill.
Discomfort in Making Judgments:	Remember that you are rating the paper, product, or performance, not the student. This is just one performance assessment—not a measure of overall ability.
The Sympathy Score:	"The student was really trying," "...seems to be such a nice kid," "...chose a hard topic," "...had a tough day," etc.

—Adapted from Culham and Spandel (1993)

Chapter 3



Rubrics, Scoring Guides, and Performance Criteria

Judith Arter
Assessment Training Institute, Portland, Oregon

In Chapter 2, you learned that an analytic rubric, which allows a product or performance to be rated on multiple characteristics, provides the kind of detailed information that will help you communicate with students and plan and improve instruction. The holistic rubric, which provides only a single scale for a global rating, lacks that kind of precision but offers increased efficiency and is more commonly used in large-scale assessments, such as those administered by state departments of education.

As Carol Perlman noted, you can borrow, tweak, or create your own rubrics for classroom assessment. This chapter will help you assess the quality of your rubric before you get to the piloting stage. Read on to find out how to evaluate your rubrics for content, clarity, practicality, and technical soundness. See how rubrics can be teaching tools as well as assessment tools and learn about the evidence that links scoring rubrics to positive educational outcomes.

Rubrics, scoring guides, and performance criteria describe what to look for in products or performances to judge their quality. There are essentially two uses for rubrics in the classroom:

- ✓ To gather information on students in order to plan instruction, track student progress toward important learning targets, and report progress to others.

This chapter was adapted, with permission of the author, from a paper representated at the American Educational Research Association annual meeting in New Orleans in 2000. It presents ideas developed at greater length in the book *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance* by Judy Arter and Jay McTighe. © 2001 Corwin Press. The Web site for the Assessment Training Institute is http://www.assessmentinst.com.

- ✓ To help students become increasingly proficient on the very performances and products also being assessed. In other words, rubrics provide criteria that can be used to enhance the quality of student performance, not simply evaluate it.

The idea behind the first point is that rubrics, scoring guides, and performance criteria help define important outcomes for students. Many times, some of the more complex outcomes for students are not very well defined in our minds. What is critical thinking, life-long learning, or communication in mathematics, for example, and how will we know when students do these things adequately? Well-crafted rubrics help us define these learning targets so we can plan instruction more effectively, be more consistent in scoring student work, and be more systematic in reporting student progress.

The idea behind the second point is that when students know the criteria for quality in advance of their performance, they are provided with clear goals for their work. They don't have to guess about what is most important or how their performance will be judged. Further, students learn these criteria and can use them over and over again, deepening their understanding of quality with time. George Hillocks (1986) stated it well:

> *Scales, criteria, and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new material even when they do not have the criteria in front of them.*

There is tantalizing evidence that using criteria in these two ways has an impact on teaching and student achievement (Arter, et al., 1994; Borko et al., 1997; Clarke and Stephens, 1996; Khattri, 1995; Hillocks, 1986; OERI, 1997). The use of rubrics to enhance student learning is a specific (and classic) example of the more general principle of student involvement in their own assessment. Student involvement can take various forms—developing and using assessments, self-assessment, tracking one's own progress, or communicating about one's progress and success (Stiggins, 1999a; Black and Wiliam, 1998). In addition to the specific evidence that student use of rubrics can improve achievement, there is more general evidence that improved classroom assessment, including better quality teacher-made assessments, better communication about achievement, and student involvement in their own assessment, can improve student achievement, sometimes dramatically (Black and Wiliam, 1998; Crooks, 1988).

However, rubrics aren't magic. Not any old rubric, used in any old way, will necessarily have positive effects on student learning. If a rubric doesn't include the features of work that really define quality, we will teach to the wrong target and students will learn to the wrong target. If criteria are not clearly stated in a rubric, they will not be much good in illuminating the nature of quality. Rubrics must be of high quality in order to have positive effects in the classroom. So how can we judge the quality of a rubric?

You and your colleagues can start the process of evaluating the rubric(s) you're considering by asking yourselves two questions:

1. What do we, as teachers, want rubrics to do for us in the classroom?
2. What features do rubrics need in order for them to accomplish these things?

Some common responses to the first question are clarify learning expectations, help students evaluate themselves, help plan instruction, communicate with parents and students, and/or encourage consistent scoring across students and assignments. Some of the features that might help your rubrics serve these purposes are clearly defined terms, various levels of performance, positive, student-friendly language, etc.

Next look at a sample rubric and see if it calls to mind anything else you would like to add to your list of features of quality. Then compare your list to the Metarubric Summary in Figure 1. What are the matches?

Figure 1: Metarubric Summary

Metarubric = Criteria for judging the quality of rubrics—a rubric for rubrics

1. Content: What counts?
   - ✓ Does it cover everything of importance—doesn't leave important things out?
   - ✓ Does it leave out unimportant things? What they see is what you'll get.
2. Clarity: Does everyone understand what is meant?
   - ✓ How easy is it to understand what is meant?
   - ✓ Are terms defined?
   - ✓ Are various levels of quality defined?
   - ✓ Are there samples of work to illustrate levels of quality?
3. Practicality: Is it easy to use by teachers and students?
   - ✓ Would students understand what is meant?
   - ✓ Could students use it to self-assess?
   - ✓ Is the information provided useful for planning?
   - ✓ Is the rubric manageable?
4. Technical Quality/Fairness: Is it reliable and valid?
   - ✓ Is it reliable? Would different raters give the same scores?
   - ✓ Is it valid? Do the ratings actually represent what students can do?
   - ✓ Is it fair? Does the language in the rubric adequately describe quality for all students? Are there racial, cultural, gender biases?

Figure 2 expands the Metarubric Summary by providing more details about the four traits of content/coverage, clarity, practicality, and technical soundness/fairness. After you've had a chance to work with the Metarubric Summary in Figure 1, you may wish to underline key phrases under each trait in Figure 2 to guide you in evaluating real rubrics. As you practice applying the metarubric to rubrics you're considering for classroom use, you'll gain a deeper understanding of the performance criterion.

Figure 2: Criteria for Assessing Criteria (Metarubric)

This is a rubric for evaluating the quality of rubrics—a metarubric. This metarubric is intended to help users understand the features that make rubrics, scoring guides, and performance criteria high quality for use in the classroom as assessment and instructional tools.

It was developed for classroom assessments, not large-scale assessments. Although many of the features of quality would be the same for both uses, large-scale assessments might require features in rubrics that would be counter-productive in a rubric intended for classroom use.

Please note that we are using the terms *rubric*, *scoring guide*, and *performance criteria* to mean the same thing—written statements of the characteristics and/or features that lead to a high-quality product or performance on the part of students.

The descriptors under each trait in the metarubric are not meant as a checklist. Rather, they are meant as indicators that help the user focus on the correct *level of quality* of the rubric under consideration. An odd number of levels is used because the middle level ("on its way") represents a balance of strengths and weaknesses—the rubric under consideration is strong in some ways, but weak in others. A strong score ("ready for use") doesn't necessarily mean that the rubric under consideration is perfect; rather, it means that one would have very little work to do to get it ready for use. A weak score ("not ready for prime time") means that the rubric under consideration needs so much work that it probably isn't worth the effort—it's time to find another one. It might even be easier to begin from scratch. A middle level score means that the rubric under consideration is about halfway there—it would take some work to make it usable, but it probably is worth the effort.

Additionally, a middle score does not mean "average." This is a criterion-reference scale, not a norm-referenced one. It is meant to describe levels of quality in a rubric, not to describe what is currently available. It could be that the average rubric currently available is closer to a "needs revision" than to an "on its way."

The scale could easily be expanded to a five-point scale. In this case, think of "4" as a balance of characteristics from the "5" and "3" levels. Likewise, a "2" would be a balance of characteristics from the "3" and "1" levels.

No performance standard has been set on the metarubric. In other words, there is no "cut" score that indicates when a rubric is "good enough."

Metarubric Trait 1: Content/Coverage

The content of a rubric defines what to look for in a student's product or performance to determine its quality. Rubric content constitutes the final definition of content standards because the rubric describes what will "count." Regardless of what is stated in content standards, curriculum frameworks, or instructional materials, the content of the rubric is what teachers and students will use to determine what they need to do in order to succeed; what they see is what you'll get. Therefore, it is essential that the rubric cover all essential aspects that define quality in a product or performance and leave out all things trivial.

If a rubric contains things other than those that really distinguish quality, teachers won't buy into them, students will not learn what really contributes to quality, and one might find oneself having to score down work that one feels in one's heart is strong or give good ratings to work that one feels in one's heart is weak.

Questions to ask oneself when evaluating a rubric for content are: Can I explain why each thing I have included in my rubric is essential to a high-quality performance? Can I cite references that describe the best thinking in the field on the nature of high-quality performance? Can I describe what I left out and why I left it out? Do I ever find performances or products that are scored low (or high) that I really think are good (or bad)? (If so, it is time to reevaluate the content of the rubric.) Is this worth the time devoted to it?

BEST COPY AVAILABLE

Ready to Roll:

- ✓ There is justification for the dimensions of student performance or work that are cited as being indicators of quality. Content is based on the best thinking in the field.
- ✓ The content has the "ring of truth"—your experience as a teacher confirms that the content is truly what you do look for when you evaluate the quality of student work or performance.
- ✓ If counting the number of something (such as the number of references at the end of a research report) is included as an indicator, such counts really are indicators of quality. (Sometimes, for example, two really good references are better than 10 bad ones. Or, in writing, 10 errors in spelling all on different words is more of a problem than 10 errors in spelling all on the same word.)
- ✓ The relevant emphasis on various features of performance is right—things that are more important are stressed more; things that are less important are stressed less.
- ✓ Definitions of terms are correct—they reflect current thinking in the field.
- ✓ The number of points used in the rating scale make sense. In other words, if a five-point scale is used, is it clear why? Why not a four- or six-point scale? The level of precision is appropriate for the use.
- ✓ The developer has been selective, yet complete. There is a sense that the features of importance have been covered well, yet there is no overload.
- ✓ You are left with few questions about what was included or why it was included.
- ✓ The rubric is insightful. It really helps you organize your thinking about what it means to perform with quality. The content will help you assist students to understand the nature of quality.

On Its Way but Needs Revision:

- ✓ The rubric is about halfway there on content. Much of the content is relevant, but you can easily think of some important things that have been left out or that have been given short shrift.
- ✓ The developer is beginning to develop the relevant aspects of performance. You can see where the rubric is headed, even though some features might not ring true or are out of balance.
- ✓ Although much of the rubric seems reasonable, some of it doesn't seem to represent current best thinking about what it means to perform well on the product or skill under consideration.
- ✓ Although the content seems fairly complete, the rubric sprawls—it's not organized very well.
- ✓ Although much of the rubric covers that which is important, it also contains several irrelevant features that might lead to an incorrect conclusion about the quality of the students' performance.

Not Ready for Prime Time:

- ✓ You can think of many important dimensions of a high-quality performance or product that are not included in the rubric.
- ✓ There are several irrelevant features. You find yourself asking, "Why assess this?" or "Why should this count?" or "Why is it important that students do it this way?"
- ✓ Content is based on counting the number of something when quality is more important than quantity.

- ✓ The rubric seems "mixed up"—things that go together don't seem to be placed together. Things that are different are put together.
- ✓ The rubric is very out of balance—features of importance are weighted incorrectly. (For example a business letter might have several categories that relate to format but only one that relates to content and organization.)
- ✓ Definitions of terms are incorrect—they don't reflect current best thinking in the field.
- ✓ The rubric is an endless list of seemingly everything the developer can think of that might be even marginally important. There is no organization to it. The developer seems unable to pick out what is most significant or telling. The rubric looks like a brainstormed list.
- ✓ You are left with many questions about what was included and why it was included.
- ✓ There are many features of the rubric that might lead a rater to an incorrect conclusion about the quality of a student's performance.
- ✓ The rubric doesn't seem to align with the content standard it's supposed to assess.

Metarubric Trait 2: Clarity

A rubric is clear to the extent that teachers, students, and others are likely to interpret the statements and terms in the rubric the same way. Please notice that a rubric can be strong on the trait of content/coverage, but weak on the trait of clarity—the rubric seems to cover the important dimensions of performance, but they aren't described very well. Likewise, a rubric can be strong on the trait of clarity, but weak on the trait of content/coverage—it's very clear what the rubric means, it's just not very important stuff.

Questions to ask oneself when evaluating a rubric for clarity are: Would two teachers give the same rating on a product or performance? Can I define each statement in my rubric in such a way that students can understand what I mean? Could I find examples of student work or performances that illustrate each level of quality? Would I know what to say if a student asks, "Why did I get this score?"

Ready to Roll:

- ✓ The rubric is so clear that different teachers would give the same rating to the same product or performance.
- ✓ A single teacher could use the rubric to provide consistent ratings across assignments, time, and students.
- ✓ Words are specific and accurate. It is easy to understand just what is meant.
- ✓ There are several samples of student products or performance that illustrate each score point. It is clear why each sample was scored the way it was.
- ✓ Terms are defined.
- ✓ There is just enough descriptive detail in the form of concrete indicators, adjectives, and descriptive phrases to allow you to match a student performance to the "right" score.
- ✓ There is not an overabundance of descriptive detail—the developer seems to have a sense of that which is most telling.
- ✓ The basis for assigning ratings or checkmarks is clear. Each score point is defined with indicators and descriptions.

On Its Way but Needs Revision:

- ✓ Major headings are defined, but there is little detail to help the rater choose the proper score points.
- ✓ There is some attempt to define terms and include descriptors, but it doesn't go far enough.
- ✓ Teachers would agree on how to rate some things in the rubric while others are not well defined and would probably result in disagreements.
- ✓ A single teacher would probably have trouble being consistent in scoring across students or assignments.

Not Ready for Prime Time:

- ✓ Language is so vague that almost anything could be meant. You find yourself saying things like: "I'm confused," or "I don't have any idea what they mean by this."
- ✓ There are no definitions for terms used in the rubric or the definitions provided don't help or are incorrect.
- ✓ The rubric is little more that a list of categories to rate followed by a rating scale. Nothing is defined. Few descriptors are given to define levels of performance.
- ✓ No sample student work is provided that illustrates what is meant.
- ✓ Teachers are unlikely to agree on ratings because there are so many different ways a descriptor can be interpreted.
- ✓ The only way to distinguish levels is through words such as "extremely," "very," "some," "little," and "none" or "completely," "substantially," "fairly well," "little," and "not at all."

Trait 3: Practicality

Having clear criteria that cover the right "stuff" means nothing if the system is too cumbersome to use. The trait of practicality refers to ease of use. Can teachers and students understand the rubric and use it easily? Does it give them the information they need for instructional decision making and tracking student progress toward important learning outcomes? Can the rubric be used as more than just a way to assess students? Can it also be used to improve the very achievement being assessed?

Ready to Roll:

- ✓ The rubric is manageable—there are not too many things to remember and teachers and students can internalize them easily.
- ✓ It is clear how to translate results into instruction. For example, if students appear to be weak in writing, is it clear what should be taught to improve performance?
- ✓ The rubric is usually analytical trait rather than holistic, when the product or skill is complex.
- ✓ The rubric is usually general rather than task-specific. In other words, the rubric is broadly applicable to the content of interest; it is not tied to any specific exercise or assignment.
- ✓ If task-specific and/or holistic rubrics are used, their justification is clear and appropriate. Justifications could include: (a) the complexity of the skill being assessed—a "big" skill would require an analytical trait rubric while a "small" skill might need only a single holistic rubric; or (b) the nature of the skill being assessed—understanding a concept might require a task-specific rubric, while

demonstrating a skill (such as an oral presentation) might imply a general rubric.

- ✓ The rubric can be used by students themselves to revise their own work, plan their own learning, and track their own progress. There is assistance on how to use the rubric in this fashion.
- ✓ There are "student-friendly" versions.
- ✓ The rubric is so clear that a student doing poorly would know exactly what to do in order to improve.
- ✓ The rubric is visually appealing to students; it draws them into its use.
- ✓ The rubric is divided into easily understandable chunks (traits) that help students grasp the essential aspects of a complex performance.
- ✓ The language used in the rubric is developmental—low scores do not imply "bad" or "failure."

On Its Way but Needs Revision:

- ✓ The rubric might provide useful information, but might not be easy to use.
- ✓ The rubric might be generic, yet holistic (when to be of maximal use, an analytical trait rubric would be better).
- ✓ The rubric has potential for teacher use, but would need some "tweaking"—combining long lists of attributes into traits or adjusting the language so it is clear what is intended.
- ✓ The rubric has potential for student self-use, but would need some "tweaking." This could include wording changes, streamlining, or making the format more appealing.
- ✓ Students could accurately rate their own work or performances, but it might not be clear to them what to do to improve.
- ✓ Although there are some problems, it would be easier to try and fix the rubric than look elsewhere.

Not Ready for Prime Time:

- ✓ There is no justification given for the type of rubric used—holistic or analytical trait; task-specific or generic. You get the feeling that the developer didn't know what options were available and just did what seemed like a good idea at the time.
- ✓ There seems to be no consideration of how the rubric might be useful to teachers. The intent seems to be only large-scale assessment efficiency.
- ✓ The rubric is not manageable—there is an overabundance of things to rate andit would take forever or everything is presented all at once and might overwhelm the user.
- ✓ It is not clear how to translate results into instruction.
- ✓ The rubric is worded in ways students would not understand.
- ✓ Fixing the rubric for student use would be harder than looking elsewhere.

Trait 4: Technical Soundness/Fairness

It is important to have "hard" evidence that the performance criteria adequately measure the goal being assessed, that they can be applied consistently, and that there is reason to believe that the ratings actually do represent what students can do. Although this might be beyond the scope of what individual classroom teachers can do, we all still have the responsibility to ask hard questions when we adopt or develop a rubric. The following are the things all educators should think about.

Ready to Roll:

- ✓ There is technical information associated with the rubric that describes rater agreement rates and the conditions under which such agreement rates can be obtained. These rater agreement rates are at least 65 percent exact agreement, and 98 percent within one point.
- ✓ The language used in the rubric is appropriate for the diversity of students found in typical classrooms. The language avoids stereotypic thinking, appeals to various learning styles, and uses language that English language learners would understand.
- ✓ There have been formal bias reviews of rubric content, studies of ratings under the various conditions in which ratings will occur, and studies that show that such factors as handwriting or gender or race of the student doesn't affect judgment.
- ✓ Wording is supportive of students—it describes the status of a performance rather than makes judgments of student worth.

On Its Way but Needs Revision:

- ✓ There is technical information associated with the rubric that describes rater agreement rates and the conditions under which such agreement rates can be obtained. These rater agreement rates aren't, however, at the levels described under "ready to roll." But, this might be due to less than adequate training of raters rather than to the scale itself.
- ✓ The language used in the rubric is inconsistent in its appropriateness for the diversity of students found in typical classrooms. But, these problems can be easily corrected.
- ✓ The authors present some hard data on the technical soundness of the rubric, but this has holes.
- ✓ Wording is inconsistently supportive of students, but could be corrected easily.

Not Ready for Prime Time:

- ✓ There is no technical information associated with the rubric.
- ✓ There have been no studies on the rubric to show that it assesses what is intended.
- ✓ The language used in the rubric is not appropriate for the diversity of students found in typical classrooms. The language might include stereotypes, appeal to some learning styles over others, and might put English language learners at a disadvantage. These problems are not easily corrected.
- ✓ The language used in the rubric might be hurtful to students. For example, at the low end of the rating scale, terms such as "clueless" or "has no idea how to proceed" are used. These problems would not be easy to correct.

©Assessment Training Institute September 2001. Used with permission.

Having high-quality rubrics to assess students and make instructional decisions is very important. The final piece of the puzzle is in place when we use rubrics to help students learn to assess themselves. The heart of academic competence is, after all, self-assessment-knowing what to look for in one's own work to decide what could be improved, and then knowing how to improve it. Figure 3 presents seven strategies for teaching students how to self-assess using rubrics, once high-quality rubrics are in place.

Figure 3: Seven Strategies for Using a Scoring Guide as a Teaching Tool

1. *Teach students the language of quality—the concepts behind strong performance.* How do your students already describe what a strong product or performance looks like? How does their prior knowledge relate to the elements (traits) of the scoring guide you will use? How can you get them to refine their vision of quality?

- ✓ Ask students to brainstorm characteristics of good-quality work.
- ✓ Show samples of work (low and high quality) and ask them to expand their list of quality features.
- ✓ Ask students if they'd like to see what teachers think. (They always want to.) Have them analyze how student-friendly versions of the scoring guide match to what they said.

2. *Read (view), score, and discuss anonymous sample products or performances.* Use some work that is strong and some that is weak; include some work representing problems they commonly experience, especially the problems that drive you nuts. Ask students to use the rubric(s) to "score" real samples of student work. Since there is no single correct score, only justifiable scores, ask students to justify their scores using wording from the rubric. Begin with a single trait. Progress to multiple traits when students are proficient with single-trait scoring.

3. *Let students use the scoring guide to practice and rehearse revising.* It's not enough merely to ask students to judge work and justify their judgments. Students also need to understand how to revise work to make it better. Begin by choosing work that needs revision on a single trait.

- ✓ Ask students to brainstorm advice for the author on how to improve his or her work. Then ask students (in pairs) to revise the work using their own advice. OR-
- ✓ Ask students to write a letter to the creator of the sample, suggesting what s/he could do to make the sample strong for the trait discussed. OR-
- ✓ Ask students to work on a product or performance of their own that is currently in process, revising for the trait discussed.

4. *Share examples of products or performances—both strong and weak—from life beyond school.* Have them analyze these samples for quality using the scoring guide.

5. *Model creating the product or performance yourself.* Show the messy underside-the true beginnings; how you think through decisions along the way. OR-Ask students to analyze your work for quality and make suggestions for improvement. Revise your work using their advice. Ask them to review it again for quality. Students love doing this.

6. *Encourage students to share what they know.* People consolidate understanding when they practice describing and articulating criteria for quality. Ask students to:

- ✓ Write self-reflections, letters to parents, and papers describing the process they went through to create a product or performance. Use the language of the scoring guide.
- ✓ Revise the scoring guide for younger students, make bulleted lists of elements of quality, develop posters illustrating the traits, or write a description of quality as they now understand it (I used to ..... but now I .....). Use the language of the scoring guide.
- ✓ Participate in conferences with parents and/or teachers to share their achievement.

7. *Design lessons and activities around the traits of the scoring guide.* Reorganize what you already teach and find or design additional lessons.

Judy Arter and Jan Chappuis. Student-Involved Performance Assessment (training video). 2001. Assessment Training Institute, Portland, OR: 800-480-3060. Reprinted with permission of the author. Adapted from work done at Northwest Regional Educational Laboratory, Portland, OR.

Beware—it can be easy to "over-rubric." We need to pick and choose the products or skills that would most benefit from this emphasis. A good rule of thumb is this: simple learning target, simple assessment; complex learning target, complex assessment. Knowledge and simple skills (e.g., long division) can be assessed well using multiple-choice, matching, true-false, and short-answer formats. Writing, mathematical problem solving, science process skills, critical thinking, oral presentations, and group collaboration skills probably need a performance assessment.

Chapter 4

Scoring Rubric Development: Validity and Reliability

Barbara M. Moskal and Jon A. Leydens
Colorado School of Mines

In the previous chapter on evaluating scoring rubrics, Judy Arter recommended that teachers consider the technical soundness and fairness of a rubric before using it to determine whether the performance criteria adequately measure the goal being addressed and whether they can be applied consistently. Rubrics should obviously not be biased against any particular group and should measure only those things that all children have had an opportunity to learn. Ideally, two raters using the same rubric would agree at least 65 percent of the time, and be within one point of each other a full 98 percent of the time. In this chapter, the technical issues of validity and reliability are explored in greater detail. Validity is discussed in terms of content-related, construct-related, criterion-related, and consequential evidence. Reliability encompasses interrater reliability—the consistency of scores assigned by two independent raters—and intrarater reliability—scores assigned by the same rater at different periods in time.

Although many teachers have been exposed to the statistical definitions of the terms validity and reliability in teacher preparation courses, these courses often do not discuss how these concepts are related to classroom practices (Stiggins, 1999b). One purpose of this chapter is to provide clear definitions of validity and reliability and illustrate these definitions through examples. A second purpose is to clarify how these issues may be addressed in the development of scoring rubrics. Scoring rubrics are descriptive scoring schemes that are developed by teachers or other evaluators to guide the analysis of the products and/or processes of students' efforts

This chapter first appeared in the online, peer-reviewed journal, *Practical Assessment, Research & Evaluation*, 7 (10), available at http://ericae.net/pare/.

BEST COPY AVAILABLE

(Brookhart, 1999; Moskal, 2000). The ideas presented here are applicable for anyone using scoring rubrics in the classroom, regardless of the discipline or grade level.

Validity

Validation is the process of accumulating evidence that supports the appropriateness of the inferences that are made of student responses for specified assessment uses. Validity refers to the degree to which the evidence supports that these interpretations are correct and that the manner in which the interpretations are used is appropriate (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Three types of evidence are commonly examined to support the validity of an assessment instrument: content, construct, and criterion. This section begins by defining these types of evidence and is followed by a discussion of how evidence of validity should be considered in the development of scoring rubrics.

Content-Related Evidence

Content-related evidence refers to the extent to which a student's responses to a given assessment instrument reflect that student's knowledge of the content area that is of interest. For example, a history exam in which the questions use complex sentence structures may unintentionally measure students' reading comprehension skills rather than their historical knowledge. A teacher who is interpreting a student's incorrect response may conclude that the student does not have the appropriate historical knowledge, when actually that student does not understand the questions. The teacher has misinterpreted the evidence—rendering the interpretation invalid.

Content-related evidence is also concerned with the extent to which the assessment instrument adequately samples the content domain. A mathematics test that primarily includes addition problems would provide inadequate evidence of a student's ability to solve subtraction, multiplication, and division problems. Correctly computing fifty addition problems and two multiplication problems does not provide convincing evidence that a student can subtract, multiply, or divide.

Content-related evidence should also be considered when developing scoring rubrics. The task shown in Figure 1 was developed by the Quantitative Understanding: Amplifying Student Achievement and Reasoning Project (Lane, et. al, 1995) and requests that the student provide an explanation. The intended content of this task is decimal density. In developing a scoring rubric, a teacher could unintentionally emphasize the nonmathematical components of the task. For example, the resultant scoring criteria might emphasize sentence structure and/or spelling at the expense of the mathematical knowledge that the student displays. The student's score, which is interpreted as an indicator of the student's mathematical knowledge, would actually be a reflection of the student's grammatical skills. Based on this scoring system, the resultant score would be an inaccurate measure of the student's mathematical knowledge. This discussion does not suggest that sentence structure and/or spelling cannot be assessed through this task. If the assessment is intended to examine sentence structure, spelling, and mathematics, then the score categories should reflect all of these areas.

Figure 1. Decimal Density Task

Dena tried to identify all the numbers between 3.4 and 3.5. Dena said, "3.41, 3.42, 3.43, 3.44, 3.45, 3.46, 3.47, 3.48 and 3.49. That's all the numbers that are between 3.4 and 3.5."

Nakisha disagreed and said that there were more numbers between 3.4 and 3.5.

A. Which girl is correct?
Answer:

B. Why do you think she is correct?

Construct-Related Evidence

Constructs are processes that are internal to an individual. An example of a construct is an individual's reasoning process. Although reasoning occurs inside a person, it may be partially displayed through results and explanations. An isolated correct answer, however, does not provide clear and convincing evidence of the nature of the individual's underlying reasoning process. Although an answer results from a student's reasoning process, a correct answer may be the outcome of incorrect reasoning. When the purpose of an assessment is to evaluate reasoning, both the product (i.e., the answer) and the process (i.e., the explanation) should be requested and examined.

Consider the problem shown in Figure 1. Part A of this problem requests that the student indicate which girl is correct. Part B requests an explanation. The intention of combining these two questions into a single task is to elicit evidence of the students' reasoning process. If a scoring rubric is used to guide the evaluation of students' responses to this task, then that rubric should contain criteria that addresses both the product and the process. An example of a holistic scoring rubric that examines both the answer and the explanation for this task is shown in Figure 2.

Figure 2. Example Rubric for Decimal Density Task

Proficient:	Answer to part A is Nakisha. Explanation clearly indicates that there are more numbers between the two given values.
Partially Proficient:	Answer to part A is Nakisha. Explanation indicates that there are a finite number of rational numbers between the two given values.
Not Proficient:	Answer to part A is Dana. Explanation indicates that all of the values between the two given values are listed.

Note: This rubric is intended as an example and was developed by the authors. It is not the original QUASAR rubric, which employs a five-point scale.

Evaluation criteria within the rubric may also be established that measure factors that are unrelated to the construct of interest. This is similar to the earlier example in which spelling errors were being examined in a mathematics assessment. However, here the concern is whether the elements of the responses being evaluated are appropriate indicators of the underlying construct. If the construct to be examined is reasoning, then spelling errors in the student's explanation are irrelevant to the purpose of the assessment and should not be included in the evaluation criteria.

On the other hand, if the purpose of the assessment is to examine spelling and reasoning, then both should be reflected in the evaluation criteria. Construct-related evidence is the evidence that supports that an assessment instrument is completely and only measuring the intended construct.

Reasoning is not the only construct that may be examined through classroom assessments. Problem solving, creativity, writing process, self-esteem, and attitudes are other constructs that a teacher may wish to examine. Regardless of the construct, an effort should be made to identify the facets of the construct that may be displayed and that would provide convincing evidence of the students' underlying processes. These facets should then be carefully considered in the development of the assessment instrument and in the establishment of scoring criteria.

Criterion-Related Evidence

The final type of evidence that will be discussed here is criterion-related evidence. This type of evidence supports the extent to which the results of an assessment correlate with a current or future event. Another way to think of criterion-related evidence is to consider the extent to which the students' performance on the given task may be generalized to other, more relevant activities (Rafilson, 1991).

A common practice in many engineering colleges is to develop a course that "mimics" the working environment of a practicing engineer (e.g., Sheppard and Jeninson, 1997; King, Parker, Grover, Gosink, and Middleton, 1999). These courses are specifically designed to provide the students with experiences in "real" working environments. Evaluations of these courses, which sometimes include the use of scoring rubrics (Leydens and Thompson, 1997; Knecht, Moskal, and Pavelich, 2000), are intended to examine how well prepared the students are to function as professional engineers. The quality of the assessment is dependent upon identifying the components of the current environment that will suggest successful performance in the professional environment. When a scoring rubric is used to evaluate performances within these courses, the scoring criteria should address the components of the assessment activity that are directly related to practices in the field. In other words, high scores on the assessment activity should suggest high performance outside the classroom or at the future work place.

Validity Concerns in Rubric Development

Concerns about the valid interpretation of assessment results should begin before the selection or development of a task or an assessment instrument. A well-designed scoring rubric cannot correct for a poorly designed assessment instrument. Since establishing validity is dependent on the purpose of the assessment, teachers should clearly state what they hope to learn about the responding students (i.e., the

purpose) and how the students will display these proficiencies (i.e., the objectives). The teacher should use the stated purpose and objectives to guide the development of the scoring rubric.

In order to ensure that an assessment instrument elicits evidence that is appropriate to the desired purpose, Hanny (2000) recommends numbering the intended objectives of a given assessment and then writing the number of the appropriate objective next to the question that addresses that objective. In this manner, any objectives that have not been addressed through the assessment will become apparent. This method for examining an assessment instrument may be modified to evaluate the appropriateness of a scoring rubric. First, clearly state the purpose and objectives of the assessment. Next, develop scoring criteria that address each objective. If one of the objectives is not represented in the score categories, then the rubric is unlikely to provide the evidence necessary to examine the given objective. If some of the scoring criteria are not related to the objectives, then, once again, the appropriateness of the assessment and the rubric is in question. This process for developing a scoring rubric is illustrated in Figure 3.

Figure 3. Evaluating the Appropriateness of Scoring Categories to a Stated Purpose

Step 1: State the assessment purpose and objectives.
Step 2: Develop score criteria for each objective.
Step 3: Reflect on the following:

1. Are all of the objectives measured through the scoring criteria?
2. Are any of the scoring criteria unrelated to the objectives?

Reflecting on the purpose and the objectives of the assessment will also suggest which forms of evidence—content, construct, and/or criterion—should be given consideration. If the intention of an assessment instrument is to elicit evidence of an individual's knowledge within a given content area, such as historical facts, then the appropriateness of the content-related evidence should be considered. If the assessment instrument is designed to measure reasoning, problem solving or other processes that are internal to the individual and, therefore, require more indirect examination, then the appropriateness of the construct-related evidence should be examined. If the purpose of the assessment instrument is to elicit evidence of how a student will perform outside of school or in a different situation, criterion-related evidence should be considered.

Being aware of the different types of evidence that support validity throughout the rubric development process is likely to improve the appropriateness of the interpretations when the scoring rubric is used. Validity evidence may also be examined after a preliminary rubric has been established. Table 1 displays a list of questions that may be useful in evaluating the appropriateness of a given scoring rubric with respect to the stated purpose. This table is divided according to the type of evidence being considered.

Table 1: Questions To Examine Each Type of Validity Evidence

Content	Construct	Criterion
Do the evaluation criteria address any extraneous content?	Are all of the important facets of the intended construct evaluated through the scoring criteria?	How do the scoring criteria measure the important components of the future or related performance?
Do the evaluation criteria of the scoring rubric address all aspects of the intended content?	Are any of the evaluation criteria irrelevant to the construct of interest?	Are there any facets of the future or related performance that are not reflected in the scoring criteria?
Is there any content addressed in the task that should be evaluated through the rubric, but is not?		How do the scoring criteria reflect competencies that would suggest success on future or related performances?
		What are the important components of the future or related performance that may be evaluated through the use of the assessment instrument?

Many assessments serve multiple purposes. For example, the problem displayed in Figure 1 was designed to measure both students' knowledge of decimal density and the reasoning process that students used to solve the problem. When multiple purposes are served by a given assessment, more than one form of evidence may need to be considered.

Another form of validity evidence that is often discussed is "consequential evidence." Consequential evidence refers to examining the consequences or uses of the assessment results. For example, a teacher may find that the application of the scoring rubric to the evaluation of male and female performances on a given task consistently results in lower evaluations for the male students. The interpretation of this result may be the male students are not as proficient within the area that is being investigated as the female students. It is possible that the identified difference is actually the result of a factor that is unrelated to the purpose of the assessment. In other words, the completion of the task may require knowledge of content or constructs that were not consistent with the original purposes. Consequential evidence refers to examining the outcomes of an assessment and using these outcomes to identify possible alternative interpretations of the assessment results (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999).

Reliability

Reliability refers to the consistency of assessment scores. For example, on a reliable test, a student would expect to attain the same score regardless of when the student completed the assessment, when the response was scored, and who scored the response. On an unreliable examination, a student's score may vary based on factors that are not related to the purpose of the assessment.

Many teachers are probably familiar with the terms *test/retest reliability, equivalent-forms reliability, split half reliability and rational equivalence reliability* (Gay, 1987). Each of these terms refers to statistical methods that are used to establish consistency of student performances within a given test or across more than one test. These types of reliability are of more concern on standardized or high-stakes testing than they are in classroom assessment. In a classroom, students' knowledge is repeatedly assessed and this allows the teacher to adjust as new insights are acquired.

The two forms of reliability that typically are considered in classroom assessment and in rubric development involve rater (or scorer) reliability. Rater reliability generally refers to the consistency of scores that are assigned by two independent raters and that are assigned by the same rater at different points in time. The former is referred to as *interrater reliability* while the latter is referred to as *intrarater reliability*.

Interrater Reliability

Interrater reliability refers to the concern that a student's score may vary from rater to rater. Students often criticize exams in which their score appears to be based on the subjective judgment of their instructor. For example, one manner in which to analyze an essay exam is to read through the students' responses and make judgments as to the quality of the students' written products. Without set criteria to guide the rating process, two independent raters may not assign the same score to a given response. Each rater has his or her own evaluation criteria. Scoring rubrics respond to this concern by formalizing the criteria at each score level. The descriptions of the score levels are used to guide the evaluation process. Although scoring rubrics do not completely eliminate variations between raters, a well-designed scoring rubric can reduce the occurrence of these discrepancies.

Intrarater Reliability

Factors that are external to the purpose of the assessment can affect the manner in which a given rater scores student responses. For example, a rater may become fatigued with the scoring process and devote less attention to the analysis over time. Certain responses may receive different scores than they would have had they been scored earlier in the evaluation. A rater's mood on the given day or knowing who a respondent is may also impact the scoring process. A correct response from a failing student may be more critically analyzed than an identical response from a student who is known to perform well. Intrarater reliability refers to each of these situations in which the scoring process of a given rater changes over time. The inconsistencies in the scoring process result from influences that are internal to the rater rather than true differences in student performances. Well-

designed scoring rubrics respond to the concern of intrarater reliability by establishing a description of the scoring criteria in advance. Throughout the scoring process, the rater should revisit the established criteria in order to ensure that consistency is maintained.

Reliability Concerns in Rubric Development

Clarifying the scoring rubric is likely to improve both interrater and intrarater reliability. A scoring rubric with well-defined score categories should assist in maintaining consistent scoring regardless of who the rater is or when the rating is completed. The following questions may be used to evaluate the clarity of a given rubric:

- ✓ Are the scoring categories well defined?
- ✓ Are the differences between the score categories clear? and
- ✓ Would two independent raters arrive at the same score for a given response based on the scoring rubric?

If the answer to any of these questions is no, then the unclear score categories should be revised.

One method of further clarifying a scoring rubric is through the use of anchor papers. Anchor papers are a set of scored responses that illustrate the nuances of the scoring rubric. A given rater may refer to the anchor papers throughout the scoring process to illuminate the differences between the score levels.

After every effort has been made to clarify the scoring categories, other teachers may be asked to use the rubric and the anchor papers to evaluate a sample set of responses. Any discrepancies between the scores that are assigned by the teachers will suggest which components of the scoring rubric require further explanation. Any differences in interpretation should be discussed, and appropriate adjustments to the scoring rubric should be negotiated. Although this negotiation process can be time-consuming, it can also greatly enhance reliability (Yancey, 1999).

Another reliability concern is the appropriateness of the given scoring rubric to the population of responding students. A scoring rubric that consistently measures the performances of one set of students may not consistently measure the performances of a different set of students. For example, if a task is embedded within a context, one population of students may be familiar with that context, and the other population may be unfamiliar with that context. The students who are unfamiliar with the given context may achieve a lower score based on their lack of knowledge of the context. If these same students had completed a different task that covered the same material but was embedded in a familiar context, their scores may have been higher. When the cause of variation in performance and the resulting scores is unrelated to the purpose of the assessment, the scores are unreliable.

Sometimes during the scoring process, teachers realize that they hold implicit criteria that are not stated in the scoring rubric. Whenever possible, the scoring rubric should be shared with the students in advance in order to allow students the opportunity to construct the response with the intention of providing convincing evidence that they have met the criteria. If the scoring rubric is shared with the students prior to the evaluation, students should not be held accountable for the unstated criteria. Identifying implicit criteria can help the teacher refine the scoring rubric for future assessments.

Concluding Remarks

Establishing reliability is a prerequisite for establishing validity (Gay, 1987). Although a valid assessment is by necessity reliable, the contrary is not true. A reliable assessment is not necessarily valid. A scoring rubric is likely to result in invalid interpretations, for example, when the scoring criteria are focused on an element of the response that is not related to the purpose of the assessment. The score criteria may be so well stated that any given response would receive the same score regardless of who the rater was or when the response was scored.

A final word of caution is necessary concerning the development of scoring rubrics. Scoring rubrics describe general, synthesized criteria that are witnessed across individual performances and therefore, cannot possibly account for the unique characteristics of every performance (Delandshere and Petrosky, 1998; Haswell and Wyche-Smith, 1994). Teachers who depend solely upon the scoring criteria during the evaluation process may be less likely to recognize inconsistencies that emerge between the observed performances and the resultant score. For example, a reliable scoring rubric may be developed and used to evaluate the performances of pre-service teachers while those individuals are providing instruction. The existence of scoring criteria may shift the rater's focus from the interpretation of an individual teacher's performances to the mere recognition of traits that appear on the rubric (Delandshere and Petrosky, 1998). A pre-service teacher who has a unique, but effective style, may acquire an invalid, low score based on the traits of the performance.

The purpose of this chapter has been to define the concepts of validity and reliability and to explain how these concepts are related to scoring rubric development. The reader may have noticed that the different types of scoring rubrics—analytic, holistic, task specific, and general—were not discussed here (for more on these, see Moskal, 2000). Neither validity nor reliability is dependent upon the type of rubric. Carefully designed analytic, holistic, task- specific, and general scoring rubrics have the potential to produce valid and reliable results.

Chapter 5

Converting Rubric Scores to Letter Grades

Northwest Regional Educational Laboratory
Portland, Oregon

We have seen in the previous chapter that a valid assessment must always be reliable, but a reliable assessment is not necessarily valid. Well-made analytical scoring rubrics can provide teachers and students with a great deal of information about aspects of student performance, including areas of strength and weakness. This information can then be used to tailor further instruction. Scoring rubrics clearly have a diagnostic purpose, but sooner or later, they also tend to feed into the grading process. How should a score on a rubic be used in a conventional grading system? Is a "3" on a five-point wholistic rubric scale equivalent to a "C"? Should points on an analytic rubric be converted to percentages and then to a letter grade? This chapter is presented as a training exercise designed to introduce teachers to the complexities of converting rubrics scores to letter grades. Read the memo, study the various grading methods proposed, and see how other teachers approach this thorny task.

Teachers want more than general discussions of issues surrounding grading and reporting. They want solutions to the situations that press them on a daily basis. One of these situations is the need to reconcile innovative student assessment designed to be standards-based and influence instruction (such as the use of rubrics to teach skills and track student performance) with the need to grade.

The example memo below was written by a district assessment coordinator in response to a request for help from teachers who were attempting to reconcile their

---

This chapter has been adapted with permission from training materials found in the Northwest Regional Educational Laboratory's publication, *Improving Classroom Assessment: A Toolkit for Professional Developers (ToolKit '98)*. The activity can help teachers and administrators explore the issues involved in converting rubric scores to grades and apply their knowledge to a specific case. For more about scoring rubrics, see http://www.nwrel.org/assessment.

use of a writing rubric for instruction with their need to grade. The teachers were using a six-point writing rubric developed by the Northwest Regional Educational Laboratory that addressed ideas, organization, voice, word choice, sentence fluency, and conventions. (This rubric has since become the 6+1 Traits(r) of Analytic Writing Assessment, with presentation included as an optional trait.) Each of the traits was scored on a scale of one to five, with five being highest. Scores were not intended to correspond to the conventional grades of "A" through "F"; further, teachers were encouraged to use the traits that made sense in any given instructional instance, weight them differently depending on the situation, and ask students to add language that made sense to them. Although the six-trait model writing rubric was used in this case study, the same issues would arise with any multi-trait rubric.

The district instituted the use of performance assessment in writing for its positive influence on instruction and its ability to track student skill levels in a useful fashion. The philosophy was this: If we clearly define what high-quality writing looks like and illustrate what we mean with samples of student writing, everyone (teachers, students, parents, etc.) will have a clearer view of the target to be hit. And, indeed, the process of determining characteristics of quality and finding samples of student work to illustrate them (and teaching students to do the same) did improve instruction and help students achieve at a higher level. The district adopted the use of the rubric scoring scheme to motivate students, provide feedback (accurate communication of achievement in relationship to standards), encourage students to self-assess, and use assessment to improve achievement.

Example Memo

February 7, 1997

To:	Teachers
From:	Linda L. Elman, Testing Coordinator Central Kitsap School District, Silverdale, Washington
RE:	Converting Rubric Scores for End-of-Quarter Letter Grades

Introduction

There is no simple or single way to manipulate rubric scores so that they can be incorporated into end-of-quarter letter grades. This paper contains a set of possible approaches. Or, you may have developed a process of your own. Whatever approach you choose to use, it is important that you inform your students about your system. How grades are calculated should be open to students rather than a mystery. In addition, you need to make sure that the process that you use is reasonable and defensible in terms of what you expect students to know and be able to do as a result of being in your class.

In all cases, you might not want to use all papers/tasks students have completed as the basis for your end-of-quarter grades-you might choose certain pieces of student work, choose to emphasize certain traits for certain pieces, let students choose their seven "best" pieces, etc. You might only want to score certain traits on certain tasks.

You might consider placing most emphasis on works completed late in the grading period. This ensures that students who are demonstrating strong achievement at the end of a term are not penalized for their early "failure." It also encourages students to take risks in the learning process. Whatever you chose to do, you need to have a clear idea in your mind how it helps you communicate how students are performing in your classroom. In the end, what you need

to have are adequate samples of student work that will allow you to be confident about how well students have mastered the skills that have been taught. (Do you have enough evidence to predict, with confidence, a student's level of mastery on his or her next piece of work?)

Down the road we will want to convene a group of teachers to come up with a common acceptable and defensible system for converting rubric scores to grades. In the short run, here are several methods that can be used to convert rubric scores to letter grades.

Methods:

The methods described here can be used with any tasks, papers, or projects that are scored using rubrics. The example used is from writing assessment, but the methods identified here are not restricted to writing.

In Table 1, we have Johnny's scores on the five pieces of writing we agreed to evaluate this term on all six traits.

Table 1: Johnny's Writing Scores on Five Papers

Johnny's Scores	Ideas and Content	Organization	Word Choice	Sentence Fluency	Voice	Conventions	Total
Paper 1	3	2	2	3	1	4	15
Paper 2	4	2	3	4	3	4	20
Paper 3	5	4	5	5	2	3	23
Paper 4	4	4	4	4	2	4	4
Paper 5	5	5	5	5	4	4	2
Totals	21	17	19	21	12	19	109

METHOD 1: Frequency of Scores Method. Develop a logic rule for assigning grades. The following are just four of many possible ways you could go about setting up a rule for assigning grades in writing.

- ✓ To get an A in writing, you have to have 50 percent of your scores at a 5, with no scores of Ideas and Content, Conventions, or Organization below 4.
- ✓ To get a B, you have to have 50 percent of your scores at a 4 or higher, with no Ideas and Content, Conventions, or Organization scores below 3, and any other score below a 3 counterbalanced by a score of 4 or higher.

OR

In this class, in writing,

- ✓ Mostly 4's and 5's is an A
- ✓ Mostly 3's and 4's is a B
- ✓ Mostly 2's and 3's is a C

OR

- ✓ To get 100 percent in writing, you have to have 50 percent of your scores at a 5, with no scores of Ideas and Content, Conventions, or Organization below 4.
- ✓ get 90 percent in writing, you have to have 50 percent of your scores at a 4 or higher, with no Ideas and Content, Convention, or Organization scores below 3, and any other score below a 3 counterbalanced by a score of 4 or higher, etc.

OR

- ✓ To get a C in writing, all writing must be at a 3 or higher. To get an A or a B, students need to choose five papers, describe the grade they should get on those papers, and justify the grade using the language of the six-trait model and specific examples from the written work.

BEST COPY AVAILABLE

Depending on how the rule finally plays out, Johnny might either get around an A (mostly 4's or 5's) or a B (lots of 4's or 5's, but there are more 4's than 5's), or 90 percent (there is one 3 in Conventions) for the writing part of his grade. Or he might get an A by citing specific examples from written work and the six-trait rubric that show he really understands what constitutes good writing and is ready to be a critical reviewer of his own work.

METHOD 2: Total Points. Add the total number of possible points students can get on rubric-scored papers.

First figure out the number of points possible. To do this, multiply the number of papers being evaluated by the number of traits assessed. Then multiply that by 5 (or the highest score possible on the rubric). In this case, with five papers we would multiple 5 (papers) times 6 (traits) times 5 (highest score on the rubric) and get a 150 total points possible.

Then we add a student's scores (Johnny has 109 points) and divide by the total possible—109 ÷ 150 = 0.73. Johnny has 73 percent of the possible points, so his writing grade will be 73 percent. We'll need to weight and combine it with other scores to come up with a single letter grade for the course.

METHOD 3: Total Weighted Points. Add the total number of possible points students can get on rubric-scored papers, weighting those traits you deem most important.

First, figure out how many points are possible. You will need to figure out which traits you are weighting. In Table 2, assume that we decided to weight Ideas and Content, Organization, and Conventions as three times more important than the other three traits. The way to come up with the total points possible is then shown. First, you add all of the scores for each trait (adding the numbers in the column), then you multiply the total in each column by its weight. Finally you add up the total in each column to come up with the grand total number of points.

Table 2: Total Possible Weighted Points

Total Possible Points	Ideas and Content	Organization	Word Choice	Sentence Fluency	Voice	Conventions	
Paper 1	5	5	5	5	5	5	
Paper 2	5	5	5	5	5	5	
Paper 3	5	5	5	5	5	5	
Paper 4	5	5	5	5	5	5	
Paper 5	5	5	5	5	5	5	
Totals	25	25	25	25	25	25	
Weights	3	3	1	1	1	3	
Weighted Total	75	75	25	25	25	75	300

BEST COPY AVAILABLE

Table 3 shows Johnny's scores again using the weighted formula. Johnny has 223 out of 300 weighted points, or 74 percent of total points in writing.

Table 3: Johnny's Writing Scores on Five Papers - WithWeighted Totals

Johnny's Scores	Ideas and Content	Organi-zation	Word Choice	Sentence Fluency	Voice	Conven-tions	
Paper 1	3	2	2	3	1	4	
Paper 2	4	2	3	4	3	4	
Paper 3	4	4	5	5	2	3	
Paper 4	5	4	4	4	2	4	
Paper 5	5	5	5	5	4	4	
Totals	21	17	19	21	12	19	
Weights	3	3	1	1	1	3	
Weighted Total	63	51	19	21	12	57	223

Depending on your focus, you might want to include only the traits you have been working on, weighting others as 0.

METHOD 4: Linear Conversion. Come up with a connection between scores on the rubrics and percents, directly. We might find that the wording of a 3 on the six-trait scales comes close to our definition of what a C is in district policy, then turn the rubric score into a percent score based on the definition. For example:

1 = 60 percent
2 = 70 percent
3 = 80 percent
4 = 90 percent
5 = 100 percent

You can then treat the rubric scores the same way you treat other grades in your grade book.

Conclusion

One difficulty with the last three approaches is that they makes the scoring and grading method seem more scientific than it is. For example, it is not always clear that the distance between a score of 1 and a score of 2 is the same as the distance between a score of 4 and a score of 5, and linear conversions, or averaging numbers, ignores those differences.

Once you, as teacher, arrive at a method for converting rubric scores to a scale that is comparable to other grades, the responsibility is on you to come up with a defensible system for weighting the pieces in the grade book to come up with a final grade for students. This part of the teaching process is part of the professional art of teaching. There is no single right way to do it; however, what is done needs to reflect evidence of **students' levels of mastery** of the targets of instructions.

BEST COPY AVAILABLE

Case Study Analysis

Note that the first and fourth methods described in the memo are essentially logic processes. They don't require adding numbers at all. Method 1 essentially says that if the student has a certain pattern of scores, he or she should get a certain grade. Several possible patterns and their relationship to grades are proposed. Method 4 asks that the teacher come up with rules for converting rubric scores to "percents." For example, an A might be equivalent to a score of 4, so 4's would get 90 percent.

The middle two methods require adding, multiplying, and dividing numbers. For example, Method 2 requires adding up all the student's earned points and comparing the sum to the total possible points (e.g., 90 percent = A). Method 3, a variation on this, weights some traits more than others before comparing earned points to total possible points.

After you read the memo, you may wish to consider or discuss with colleagues the following questions:

- ✓ What are the advantages and disadvantages of each method for converting rubric scores to grades if the purpose of grading is to accurately report student achievement status
- ✓ and support the other classroom uses of the rubric?
- ✓ Which method would you recommend and why?
- ✓ How does this compare with current practice at your school?

Here are some teachers' comments regarding the memo. What are your reactions to them?

If the main purpose of grading is accurately communicating student achievement, the best method would be 1 or 4 because dry conversion of scores to percents masks important information, like whether a student has improved over time. Why not look at only the most recent work?

If we choose different methods, the same student could get different grades. Should we have a common procedure across teachers?

Every assessment should be an opportunity to learn. Which method would best encourage this goal? Probably, the one where students choose the work to be graded.

We have to be really clear on what the grade is about, how it will be used, and what meaning it will have to students and parents. Different methods might be used at elementary and secondary.

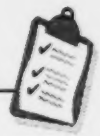
Use of rubrics in the classroom and grading come from different philosophies about the role of assessment in instruction and learning. Many times grades are just used to sort students without having a clear idea of what the grade represents—achievement, effort, helping the teacher after school, or what. If we have to grade, we need to select the method that keeps best to the spirit of the rubric.

Methods 2 and 3 can be misleading because a "3" corresponds to 60 percent, which is often seen as an "F." Yet the description for "3" doesn't seem to indicate failing work. Are we standards-based or not?

Many teachers dislike the total points and weighting methods because they

don't keep as much to the spirit of rubric use in the classroom or allow the grading to be an episode of learning. Groups tend to like methods that give them flexibility and encourage learning, such as the option to place more weight on papers produced later in the grading period or requiring students to choose the papers on which to be graded and provide a rationale for their choice.

Chapter 6



Phantom of the Rubric or Scary Stories from the Classroom

Joan James, Barb Deshler, Cleta Booth, and Jane Wade

Chapters 1 through 5 addressed basic issues related to performance assessment and scoring rubrics. The remaining chapters provide an applied view so you can see how implementation of scoring rubrics looks in real-world classrooms, how students might participate in their development, and how you might embark on a step-by-step process of rubric design and implementation. The following play, though unconventional in format, captures the peaks and valleys of three teachers' experiments with rubrics. We include it in the hopes that it will spark ideas about how you can work in collaboration with other teachers to experiment with rubrics in your classrooms. Having a small group with which to discuss challenges and victories is very helpful in supporting long-term change.

PROLOGUE—AN INVITATION

Acknowledging that life often proceeds much like a play or a "Choose Your Own Adventure" story, we invite you to read the acts and scenes of our research "play." This is a dramatic recreation of our struggles and insights. As you read, you will find that the actions of the characters, the unique setting, and the provocative research questions lead the play in a maze of different directions that are hard to predict.

To have one's research question change meaning at various steps along the way is probably only normal. Hindsight certainly shows that all our misadventures were

This chapter comes from *Collaboration for a Change: Teacher-Directed Inquiry about Performance Assessments: Reports of Five Teacher-Directed Inquiry Projects*, edited by Elizabeth Horsch, Audrey Kleinsasser, and Elizabeth Traver. Aurora, CO: Mid-Continent Regional Educational Laboratory, 1996. Used with permission.

BEST COPY AVAILABLE

necessary to come to the conclusions that now propel us to ask new questions and to be ready for acting out more plays and choosing new adventures in our teachings.

CHARACTERS

The Teachers

Barb Deshler: After obtaining her Master's of Library Science (MLS) degree, Barb was content with her job as reference librarian for the Laramie (WY) Public Library for a number of years. She decided, however, to return to college, where she achieved two additional bachelor's degrees, one in secondary education and the other in elementary education. After teaching a reading methods class at the University of Wyoming and substituting at the Wyoming Center for Teaching and Learning, Laramie (WCTL-L), Barb was hired to teach the multi-aged forth- and fifth-grade classroom. Previous experience included teaching for the Peace Corps in Kenya for two years. Barb believes that collaboration with students is essential for meaningful learning.

Joan James: Joan obtained her bachelor's degree in special education and her master's degree in curriculum and instruction, completing a master's thesis using qualitative research. As a veteran teacher of 19 years, Joan's experience includes eight years at the kindergarten level and ten years teaching special education for a variety of age levels and a variety of handicaps, including mentally impaired, emotionally disturbed, learning disabled, and autistically impaired. This past year she taught a multi-aged forth- and fifth-grade class at WCTL-L. Changing from a more traditional teacher-centered to a more child-centered approach over the years, Joan willingly tries out new theories and has a flexible, ever-changing teaching philosophy.

Cleta Booth: After her two sons were born, Cleta left teaching English to learn about child development and early childhood education. She completed American Montessori certification and a master's degree in early childhood special education. For nineteen years Cleta has taught normally developing young children and those with special needs in inclusive classrooms. Currently, Cleta teaches the half-day, played-based, pre-kindergarten class. She has recently been influenced by Howard Gardner's ideas concerning multiple intelligences (Gardner, 1993), and by several writers about the use of the project approach in early education (Katz and Chard, 1989, Gardner, 1991, and Edwards, Gandini, and Forman, 1993).

Jane Wade: Jane has taught a total of thirteen years at the pre-kindergarten through college level, including elementary grades, foreign language, and language arts. She has bachelor's degrees in both elementary education and Spanish, with certification in language arts and French. Jane has developed curriculum for pre-kindergarten through sixth-grade foreign language, and currently teaches foreign language to pre-kindergarten through ninth graders, as well as language arts at the sixth- and seventh-grade level at WCTL-L. Jane values her role as a facilitator of learning as she encourages student ownership of learning activities.

The WCTL-L Kids

The students at the WCTL-L are admitted on a first-come, first-serve basis without regard to intellectual ability, talent, or socio-economic status. The $275 per semester tuition is adjusted on a sliding scale to accommodate lower income families. Special education students are mainstreamed entirely into the regular classroom.

Pre-kindergarten: A mixed-aged class (three- to five-year-olds) of nine boys and seven girls of considerable ethnic diversity.

Fourth-Fifth Grades: Twenty-five students in each of two multi-aged classrooms, with an approximately equal mix of fourth and fifth graders and genders.

Sixth-Seventh Grades: Twenty-three students in each of two multi-aged middle school language arts classes with approximately equal numbers in each grade level, and 3/2 male/female mix. The classes were scheduled for 50 minutes per day, but students often worked in two and a half hour blocks of time during interdisciplinary units.

Supporting Researchers: Audrey Kleinsasser and Elizabeth Horsch led the statewide research group. Research teams around the state supported, encouraged, and edited one another's work.

Narrator

The collective voice of our experience and collaborative reflection.

Supporting Cast

The Parents: Many participate in the governance of the school and most are actively involved in the education of their children both at school and at home.

The Principal: Provides valuable behind-the-scenes support by actively encouraging teacher experimentation and collaboration.

University Students: Student teachers and practicum students team with classroom teachers as part of their teacher-education program.

THE SETTING

Early September 1994 at WCTL-L, located in the College of Education building on the University of Wyoming campus.

Cleta's preschool classroom is housed in the same area as the fourth-fifth grade classrooms of Barb and Joan. A tall wooden bookshelf partitions Cleta's room from Barb's. Across the hall is Joan's classroom. The sounds of learning dominate the area, as it is impossible to close off the open classrooms. Jane's classroom is located upstairs in a more traditional middle school setting.

The philosophy of the school emphasizes curriculum integration, multi-age interaction, and learning that is meaningful, hands-on, and real-life oriented. The school uses a variety of authentic evaluation tools, including student self-evaluation and portfolios. The students do not work from textbooks and are not graded using the traditional A-B-C-D-F format.

As the play opens, Barb and Joan, experienced teachers but new to this school setting, are unsure of the capabilities of the 50 fourth and fifth graders they are assigned to teach. Let the play begin!

ACT I

SCENE I

YUCK, THIS IS BAD! OR NOW WHAT?!?!

(Barb, Joan, and their student teachers decided to have their students engage in a nature study project to supplement the outdoor education camping experience at the end of September.)

JOAN: Hey, I know, let's put out all of our nature books and magazines and let the kids browse to find a topic that interests them.

BARB: Then we can come up with a list of project ideas for them to choose from, like constructing a poster with captions, writing and illustrating a children's book about their topic, creating a diorama, or scripting and performing a play.

JOAN: I think we should require both written and visual aspects in their projects.

BARB: Since we have four teachers, we could divide the kids into four groups. Each teacher would be in charge of conducting mini-lessons to model what is expected. Let's each come up with some project ideas, type them up, and introduce the project to the kids tomorrow.

NARRATOR: ***Thus the team haphazardly embarked on its first project. They gave the kids an hour a day each day for three weeks to complete their project, and told them they would each do a stand-up presentation of their project at the end of that time. The teachers' mini-lessons consisted of introducing the project and the variety of formats students could use, as well as modeling how to locate useful information and how to utilize it in their projects without being sanctioned for plagiarism.***

The kids were so excited that they wanted to start working right away, making it almost impossible to interest them in the teacher-presented mini-lessons. Most of the kids made quick decisions about what they wanted to do. They wasted little time researching or reading about their topic. Instead, they enthusiastically dug into creating the hands-on portion of their project in an attempt to make the visual match up to the ideal they imagined.

After two or three days of work using paint, construction paper, clay, cardboard, and scissors to construct the visual portion of their project, many of the kids began to lose interest. Their projects lacked planning and weren't turning out as well as they'd imagined. Other students couldn't seem to get motivated to put much effort at all into the project and spent a lot of the project time sitting and staring into space. Others had learned to play the school game. They had quickly (within the first two or three days) produced a grossly inadequate written and visual portion for their project and had happily announced that they were done, expecting to be allowed to have free time for the remaining two plus weeks. When encouraged to revise, edit, and extend their projects, they had absolutely no interest and put most of their energy into passive resistance. As the three weeks wore on, the four teachers spent increasing amounts of time playing prison guard as they policed the easily accessible hall and secluded corners of the classroom for off-task students. There actually were a few, well maybe a couple, students who

really got into their project and worked diligently during each project time in an effort to learn all they could and produce a high-quality product.

JOAN: *(to the group of gathered students)*: You will have two more days to complete your project. On Friday we will begin the presentations!

Panic and Pandemonium

NARRATOR: ***Those students (a lot of them!) who had wasted the majority of the three weeks started frantically slopping together their projects. Many wrote a rough paragraph about what they had learned to fulfill the written requirement. They were ready!***

(Friday. A microphone in center stage. Seeing the microphone as they enter, kids panic. Eyes bulge. They gulp. This is for real!)

NARRATOR: ***One by one the students made their way to the front of the class and read their reports into the microphone. Modeling after one another, most read as fast as they could, holding their reports in such a way as to hide their faces from their audience and peers. When finished, they quickly held up their visuals and then stood with hands by their sides and heads down awaiting the questions from their peers. Since many had done little or no research, they were unable to answer these questions factually. Feeling pressured to perform well, they simply made up answers. Many of the written reports they turned in were pitiful, containing many punctuation and spelling errors, poor sentence and paragraph structure, and holding little interest for a reader.***

The audience wasn't much better! They felt comfortable talking with neighbors or engaging in unrelated tasks during presentations and often asked silly questions in an attempt to get a laugh from the crowd.
Following the presentations, Barb, Joan, and the student teachers sat down together.

ALL FOUR TEACHERS: Phewwwwwwwwww! *(wiping furrowed brows).*

NARRATOR: ***They had a lot of work to do! These initial presentations showed the teachers how little their students knew about high-quality work and oral presentations. The students seemed initially to be intrinsically motivated but obviously lacked the tools to translate their enthusiasm into high-quality practice. This created a dilemma for the teachers. "How do we teach and motivate students to delve deeply into a topic, to plan and work diligently toward the production of a high-quality project, and to present their knowledge to an audience in an interesting and exciting way?" and "How do we communicate expectations for high-quality work to parents?" which, in essence, became this pair's research questions (but this is getting ahead of the story).***

SCENE II

WOW, MUCH BETTER!

(After school the next day.)

BARB: You know, I know a little bit about rubrics, and I think they just might be the ticket in helping us communicate to the students and their parents our perceptions of a high-quality project. They might be a good evaluation tool, too.

AMY: *(Student Teacher)*: I don't get it. What do you mean by rubrics?

BARB: Rubrics are an evaluation tool we could use to briefly, but specifically, describe high-, middle-, and low-quality projects. Then, following a student's presentation, we could evaluate by circling the specific statements that describe their project.

NARRATOR: ***This is the definition of rubric that Barb introduced: "Rubrics provide criteria that describe student performance at various levels of proficiency"*** *(Making Assessment Meaningful,* August 1994).

JOAN: Well, we obviously have to do something, and this rubric business sounds like it's worth a try.

(Barb's room, the next day after school.)

NARRATOR: ***Barb and Joan and their student teachers met for hours to collaborate in the making of this first rubric.***

JOAN: Let's thoroughly explain this rubric to the kids so they know without question what our expectations are for this next project. As Arter says, "...there are only two choices: we can either make our criteria crystal clear to students or we can make them guess" (Arter, 1993).

BARB: Let's also send a copy of the rubric home in the mail with a letter detailing the project to be sure that the parents get the information. In the letter we can instruct the kids to explain the project and the rubric thoroughly to their parents.

JOAN: Maybe we ought to have this first project done completely at home with the assistance of the parents so that both the parents and the students will have a rock-solid understanding of our expectations.

BARB: That way the parents can take a major role in guiding their children toward quality work.

NARRATOR: ***The rubric and a letter of explanation were sent home in the mail with the due date for the project a month away. Once again, the students were required to have both a written and visual portion to their project. In addition, the teachers encouraged the students to involve the audience actively in their presentations in order to keep the audience's attention.***

Presentations

(Late October. Presentation Day. Stage with microphone. Kids enter looking nervously at microphone, but take their seats quietly. One by one they present their projects.)

NARRATOR: ***After about ten presentations, the teachers sent the students off to gym and sat down to discuss what they'd seen.***

JOAN: Wow! I can't believe what I'm seeing!

BARB: These are fantastic!

AMY: Much better than those raunchy nature projects!

NARRATOR: ***After school, the team got together to discuss and compare their rubric evaluations. Remarkably, they had usually agreed in their evaluations, circling similar criteria statements and noting similar strengths and weaknesses. The written reports were of much higher quality than those seen for the nature projects. It was evident that time and effort had been taken to revise and edit. Most of the children wrote factual reports describing the wealth of knowledge they had obviously acquired through persistent research using relevant resource materials. Many of the reports were written from the viewpoint of a New World explorer or a child who lived in colonial times to make the historical topics come alive for the audience. For visuals the students created costumes, engaged their families or friends as characters in plays they had written, constructed dioramas, drew pictures and maps, made toys, or shared food representative of the colonial era.***

The teachers involved the students in a narrative self-evaluation following their presentations. They noticed that most students had a much better understanding of the criteria necessary to produce a high-quality project and presentation than they had shown on the nature projects.

While being generally pleased with the presentations and the knowledge the students gained, the teachers found that there were still some areas for student improvement. For instance, they wanted the students to learn their topics well enough to "tell" their presentation rather than read it, speaking in a clear, strong voice, and giving their audience confident eye contact. There were also a few students yet to be motivated.

ACT II

SCENE I

TEACHERS DOING RESEARCH? WHAT A WILD IDEA!

(Late October. Wyoming Interdisciplinary Conference, Casper. A session on collaborative teacher research.)

TEACHER/PRESENTER #1: It's been an exciting journey for me, a classroom teacher, to take a critical look at my own teaching and the learning of my students through a qualitative research project.

TEACHER/PRESENTER #2: Collaborating over compressed video with teachers around the state who, like me, are interested in improving the teaching and learning in our own classrooms has been both encouraging and insightful.

NARRATOR: ***It wasn't long before Barb and Joan were hooked. They wanted to be part of this relatively new model where teachers became their own in-class researchers. Back in Laramie their excitement was contagious and, before they knew it, Cleta and Jane decided to join the team.***

Shortly after Christmas, the statewide research group met in person and later continued its collaborative work via monthly interactive compressed video discussion sessions. There were a lot of heady decisions to be made at the beginning, number one being, What should our research topic be? Joan and Barb expressed the excitement that the rubrics had produced in their classrooms. Cleta was doubtful but thought it would be interesting to see if rubric criteria would be developmentally appropriate for use with pre-kindergarten children. Jane, seeking a method to help motivate her hormone-driven middle schoolers, was willing to try anything.

SCENE II

CHOOSE YOUR OWN ADVENTURE

Because the three aspects of the rubric research (Cleta at the pre-kindergarten level, Barb and Joan at the fourth- and fifth-grade level, and Jane at the middle school level) are so different, we will now allow you to choose your own adventure and delve into scenes of your choice.

For pre-kindergarten adventures, go to Scene II-A (below)
For fourth-fifth adventures, go to Scene II-B (p. 55)
For sixth-seventh adventures, go to Scene II-C (p. 59)

SCENE II-A

ADVENTURES WITH PRE-K

(January. Barb's classroom after school.)

CLETA: I hate to interrupt, but I've been thinking about my part of the teacher-researcher project. I can't figure out my research question. I don't even know if preschool kids think in terms of good work and less good work.

BARB *(grinning)*: Well, you could investigate that. Or try using a rubric, and if it doesn't work, in the worst-case scenario, you can report that... and what you learned in the process.

CLETA: I'm already working on including kids more in planning. I'm experimenting with making Know-Wonder-Learn charts when we start a unit and with asking, How can we find out? and *What shall we do?* Maybe that's all related somehow and could be part of my research.

BARB: That's similar to what we've been trying to do, to give kids more ownership of their learning. But Cleta, one thing bothers me. Would you really tell a four-year-old that something she made doesn't meet the standard of a rubric? That feels all wrong to me.

CLETA *(laughing)*: Of course not! I might point out, "Sandy, I see your block building keeps falling down and you look upset. Is there something you can change to make the bottom wider?" To me the issue isn't grading... an A, B, or C in block building is a ridiculous idea. I want to give Sandy feedback useful for making other

buildings. I wouldn't walk by and say, "Good job. Keep trying," when Sandy knows it's not a good job because it keeps falling down. That's just frustrating. It's not honest, and it isn't helpful. Besides, I'm not thinking about using rubrics to evaluate any one child's work. That could be threatening. For a first step, I'm wondering if the children and I can even evaluate one class project together, decide what makes it good, then try to apply those standards to some other class project. I'm not sure my kids can think that way.

BARB: Well, I feel better about that. How are you going to do it?

CLETA: I guess I'll dig out the tape recorder and then transcribe some of our class discussions when we finish the next project.

BARB: You ought to videotape some, too. And don't forget to tape some of your planning sessions while you are at it.

CLETA: With kids the age I teach, I think it's critical to have the parents involved. I guess I need to find out how they would judge a project. Come to think of it, that's getting ahead of myself. I don't guess I've ever even asked parents what kind of feedback they want about what their child does at school. I've always assumed I knew! I have a feeling my teacher-researcher question is getting out of hand. I'll have to explore kids' and parents' ideas!

NARRATOR: ***Cleta continued to struggle throughout the research project to make her work appropriate and meaningful for the age group but also to find meaningful connections in it to what the teachers of older students were researching.***

Ask the Parents

(Spring parent-teacher-child conference in the pre-kindergarten room. The Porter family (a pseudonym and composite of several interviewed) is gathered around a table. They have just finished viewing four-year-old Sandy's portfolio.)

CLETA: Do you have a favorite page in your journal that you'd like to share?

SANDY *(opening the journal at random and pointing to a page covered with circular multi-color scribbles)*: This one.

CLETA: Why do you like that one best?

SANDY: Because that's the day I got my new pencil that draws all colors.

CLETA: I remember that you really liked that pencil. *(Pause)* Could you find the page when we were studying castles and you made a picture that was lots different? Do you remember? And can you tell your mom and dad about it?

SANDY *(turning to page with a simple outline on it)*: That's the well in the castle. That's where the water is, and there's the bucket, and there's the rope, and it goes up here to the pulley, and here is where you pull to get the water.

CLETA *(to the parents)*: Can you see why I think this is so remarkable? The new-pencil picture was typical of Sandy's drawings up to this point. The well picture is a cross-section drawing that shows all the parts and how it works. That's pretty unusual for a four- year-old. And after this picture Sandy began using drawing to represent other ideas. Back here there's a castle, and here's an airplane. (turning again to Sandy) Sandy, why do you think this drawing of the well is a good drawing?

SANDY *(smiling hesitantly)*: Because I liked the well.

BEST COPY AVAILABLE

NARRATOR: ***That line of reasoning going no further, Cleta allowed Sandy to find activities elsewhere while she and the Porters finished the conference. Ending a few minutes before the next family arrived, Cleta explained her team teacher-research project and asked them to answer two questions to be tape-recorded.***

CLETA: First, what kind of information or feedback about your child do you want from a report or parent conference? What is helpful?

MRS. P: I guess I most want to know how my child gets along with the others. And then I like hearing about interests and activities. I wish I could slip in and watch from a one-way-mirror. I also like to know what you are studying so we can do activities at home to encourage an interest. Your weekly newsletter really helps with that.

MR. P *(interrupting)*: For me, it's a matter of, Does my child fulfill your expectations? I don't have expectations. I assign you the job of having those expectations-and letting me know what they are and how well Sandy is meeting them.

CLETA: Hmm. I'd like to talk about that again when we have more time. Right now I'd like your thoughts on my second question, What kind of evidence do you see at home about your child's learning?

MRS. P: Sandy is pretty quiet-she doesn't often tell us directly about what is happening at school, but I always know what the class is studying. New topics like castles or rocket ships just pop into the conversation, and I know it's from what is happening at school.

MR. P: And when we go to the library, Sandy always wants to get a book on whatever topic you are studying at school.

MRS. P: Sometimes I even hear the new interests and new information come up in fantasy play when our neighbor's child comes over to visit. Or Sandy will burst into a snatch of song about a rocket, or draw a picture of planets, or tell a story about a castle. It has happened with almost every topic you've studied.

CLETA: So you both do feel that you can tell that Sandy is learning-and what Sandy is learning?

BOTH PARENTS: Oh, definitely.

MR. P *(rising to leave)*: But we do need to know if Sandy is meeting your expectations.

NARRATOR: ***From learning about Joan and Barb's experience, Cleta realized that rubrics might do exactly what Mr. Porter asked. But did rubrics make sense for looking at the work of young children? She had serious doubts. It was time to talk with the children.***

Ask the Kids-1

(Early May. Fifteen three- through five-year-olds are gathered on the rug for a discussion with the teacher.)

CLETA: It's getting near the end of the year. I wonder what projects we've done this year that you think are really good projects? What did you like best that we've done? And can you tell me why you think that one's a good project?

NATHAN *(age 5)*: Rockets! We could go in it. And I like rockets.

AURORA *(age 5)*: Castles was best. We got to go in it, 'cause I like castles.

NORA *(age 5)*: Castles. I really like being the princess... no, the prince.

TOD *(age 5)*: Airplanes. I got to be the pilot.

CODY *(age 5)*: I like bananas best. Bananas are good to eat, and I liked to pretend going to the beach. And I liked the real sand.

EVAN *(age 4)*: And the boats had water and fish.

DARLA *(age 3)*: I liked to color. Housekeeping.

ROSIE *(age 5)*: Studying about Chinese things was best because I like to pretend writing Chinese and make Chinese paintings.

NARRATOR: ***The discussion continued, dominated by the older children, until almost every project had been named. Many children named several that they liked best. Cleta transcribed the tapes and looked for patterns in their responses. Meanwhile...***

Ask the Kids-2

(Same as above, two days later.)

CLETA: When we talked about your favorite projects, I was surprised no one mentioned the chocolate friendship cake. Maybe it was too long ago.

EVAN: *(age 4)*: It was gooodyyy!

NED *(age 5)*: I got to take some home.

CLETA: What made it a good project?

ALL *(in chorus)*: I brought... *(listing the ingredients.)*

CLETA: You all contributed ingredients that made it taste good. Do you mean that one thing that made it a good project is that everybody contributed? *(Many nods.)*

NARRATOR: ***Clearly each child had meant that his or her ingredient made the cake taste good. Cleta realized she was trying to lead them to a more abstract definition of a good project (her definition!). She reminded the class that they had had to make many decisions on the cake project, and the children agreed and were able to recall them. They also recalled sharing it with the the other classes and their families. With the children's agreement, Cleta formulated criteria for a good project: everyone contributes, it involves shared decision-making, and the final results are shared with others. She then asked them to apply the same criteria to evaluate another project.***

CLETA: Let's see if we can use those ideas to see if the castle project was good for the same reasons. Did everyone help make it or do it?

ALL: Yes, painted it-marked lines on it.

NARRATOR: ***With the teacher scaffolding, they could apply the set of criteria. They easily recalled decisions they had made as a group to set rules for the castle and that they had shared by giving everyone the secret password for the drawbridge, including the older children who came for weekly integration activities. Cleta realized, however, that they could not apply criteria without a great deal of adult help.***

One More Dialogue

(The house of Rosie, a very verbal, unusually reflective five-year-old, and her mother, Anne, who had volunteered in the classroom weekly.)

CLETA: Rosie, we've put all the books you made here on the table. Do you look at them and think, "This is my best book?" or "This book is good because of the pictures, and this other book is good because of the story, but this other book isn't as good because I was in a hurry?" or anything like that?

ROSIE *(emphatically)*: I don't think it's that way, because I've already made lots of books before I start.

CLETA: So they are all good books in their own way? *(Rosie nods)*. Well then, would you choose one you especially like?

(Rosie reaches for one with her name and the number 5 on the cover. It was made soon after her fifth birthday and laminated for the class to keep and to remember her when she was in kindergarten.)

CLETA: Why did you pick this one?

ROSIE: The rocket one. It's fun, and you can read it on your bed.

NARRATOR: ***Cleta probed, and Rosie offered many reasons, including "It get's done soon," "It looks like a real rocket," and "It's for you," but Cleta could recognize no coherent pattern to her responses. Cleta decided to check Rosie's previous response about projects.***

CLETA: Rosie, tell me about which was your favorite project, or the one you thought was the best project the class did this year.

ROSIE: Space! I liked pretending to land on the moon.

CLETA: Any other reasons you think space was the best project?

ROSIE: You could go inside the rocket.

NARRATOR: ***This response was interesting, because in class Rosie had picked a different project. Her answer today was at least in part because she had just been talking about her book on rockets. It also confirmed conclusions Cleta had begun to draw about what children value in projects. Rosie left to play, and parent and teacher continued the conversation.***

CLETA: Anne, you were in the class every week all year and probably have a better view of what went on than anyone. Which project do you think was the best? What do you think makes a good project for this age group?

ANNE: I think the best projects were ones that had a real personal connection and effect-like tying in what they were studying with your trip to China. They really felt connected to you while you were away. And of course castles, because children this age are all caught up in fairy tales. Space wasn't personal to them, but like dinosaurs, it's one of those topics that always seem to catch their interest. I'd say projects that involved making something or building something-something concrete-are best. Also projects that help them understand how something happens.

CLETA: Anything else?

ANNE: Parents will have a different set of issues and values from kids. Personally, I think the most important criterion for a good project is that it generates a spark of

interest that continues at home-and maybe creates an interest that lasts into the future-a sense of possibilities for learning and excitement. Rosie still talks about wanting to be an astronaut.

CLETA: I've explained what rubrics are. Do yo think some version of a rubric for our projects would be helpful to parents?

ANNE: What would be helpful is to know the individual child's contribution to a project. And a way to alert the parents to things to listen for in the children's play and fantasy and storytelling so they can build on the ideas at home that they are learning at school. I'm not interested in any kind of rating or judging of the children's work.

Touching Base

(Barb's room after school. Joan and Barb are talking. Cleta joins them)

CLETA: I'm beginning to pull together my data from all the different sources for the teacher-researcher project. I guess what I've really been doing is comparing what teachers (at least this teacher), students, and parents value in a project-defining our implied rubrics.

BARB: And...?

CLETA: My own criteria for a good group project focused on the process of it: that it requires joint problem-solving, that it encourages cooperation and joint participation *(everyone helps)*, and that it involves sharing with an audience or bringing in another group of participants. The kids could understand and make judgments about those characteristics, but they were clearly my criteria, not theirs.

Parents looked more at the effects: that it involves the children personally, shows them concretely how something is done or made, and most importantly, that it sparks an interest that encourages the child to keep learning about the subject outside of school.

The kids, of course, had their own priorities: a good project must have opportunities for fantasy and role-playing. It's even better if it creates a physical environment you can actually go into for that role-playing. And it's good if it offers opportunities for creative art work and construction. For projects that aren't the role-playing kind, the most important criterion is that it tastes good!

JOAN *(laughing)*: That's my kind!

CLETA: I've also learned more about how kids think from this research. The three-year-olds had almost nothing to say. One girl said she liked housekeeping and coloring pictures best. That's really the same criteria of fantasy play for a go-into environment with the opportunity for creative art work, but for her the larger content framework of the particular project wasn't important. The fours and fives did have interests in particular topics, like space or castles. The fives tended to talk more, to be eager to tell their favorite project and the reasons. I think they have a clearer idea of "reasons."

BARB: Didn't you tell me you were reading something that said three-year-olds thought in a completely different way?

CLETA: Yes, Janet Astington was reviewing research that shows that children begin to be able to think and talk about their own knowing at about age four. They can talk

about wanting even earlier. Maybe that's why the threes didn't have much to say. But Astington (1993) does say that children who have lots of experience hearing talk about knowing and about reasons reach the stage of being able to talk about knowing and reasons sooner. Even if I'm asking something of the group that three-year-olds can't do, maybe it models a kind of thinking that is useful for them to hear. It's still a stretch for many of the four-year-olds, but more comfortable for the fives. I did notice that when I asked even the fours and fives to pick out their best drawing from their portfolios or identify their best work, they usually seemed to pick at random. I don't think "best" is very meaningful to them.

JOAN: So, did you decide that rubrics weren't appropriate for preschoolers?

CLETA: I don't think rubrics, as you use them in the upper grades, make any sense for preschoolers. But I'm convinced that the communication of clear expectations to children and parents is important at all ages. In setting those expectations, I just want to be sure to take into account the parents' and children's values too.

NARRATOR: ***Cleta went on to tell about her concern that most American early-childhood teachers have low expectations about what young children can do. She described the art work she had seen when she visited China, and how it was two to four years ahead of what children in her class do. She also told about similar work she'd read about from preschools in Reggio Emilia, Italy (Edwards et al., 1994).***

JOAN: They don't use rubrics in China or Italy, do they?

CLETA: No, but they do hold up good work and talk about why it's good.

BARB: But doesn't everyone just copy it?

CLETA: Sure, some do, or they try to use the particular technique that was pointed out, but isn't that a useful way to learn-to be shown the characteristics of a good model? Didn't you discover that children model from each other anyway?

JOAN: They certainly do. And we found from our experience that we needed to raise our expectations and give children better feedback.

CLETA: In Reggio Emilia they use a novel approach to feedback that they call documentation. They constantly photograph and video and tape-record the children as they work. Then they make beautiful displays of all of the stages of a project as it unfolds. Children are encouraged to keep going back and adding to their work, reflecting on it, trying it in a new medium. Feedback is continuous, and it comes from the teachers, the environment, and even the other children. They don't use the specific criteria of rubrics, but they provide the feedback for children to meet the expectations for high-quality work. Early childhood teachers here should probably be doing that too.

NARRATOR: ***Cleta was still uncertain about the best way to give children and families feedback, but she was very pleased that her findings connected with those of the older-grade teachers. This collaborative research project had made her feel a more integral part of the school than anything else in her seven years there.***

For the fourth-fifth adventure, go to Scene II-B (below)
For the sixth-seventh adventure, go to Scene II-C (p. 59)
For the conclusion, go to The Rest of the Story (p. 60)

SCENE II-B

JOANIE AND BARB'S RUBRIC RUCKUS

(Early spring semester, Joan's room after school.)

JOAN: Basically, for our research question, we've decided to explore this rubric phenomenon further to see if it continues to be a successful tool for helping our students achieve high-quality work. How can we take a further step with the rubric so we continue encouraging high-quality work?

BARB: I'm not quite sure yet. Let's both think about it.

NARRATOR: ***The rubric had helped Joan and Barb communicate their expectations and most of the students had shown that they knew how to use the criteria in the rubric to produce high-quality work. Parent enthusiasm for the previous project had been high. Many of them had come in to watch their own child's presentation and then had stayed to watch others. Joan and Barb thought it was a good idea to display all the students' work, so the hall turned into a Colonial Museum. More parents, other students, and teachers in their school came to look at the projects, and comments like these were overheard: "I am just amazed by what these kids did. I have my college students do similar projects," "I wish we could do these kinds of projects at my school" (a visiting student), and from the younger and older kids in their school, "cool," and "awesome."***

Some parents made it a point to say how much they had appreciated the letter and the rubric that had been sent home. "It helped me understand exactly what my child needed to do for a good project. It gave us excellent guidance," said more than one parent. It was, however, time to move on.

(A few days later in the hall between Joan's and Barb's rooms.)

JOAN: I think I've got it! For the next rubric adventure let's give the students the opportunity to explore something they are interested in. This could be their chance to explore something they're curious about, or even to share something they already know a lot about.

BARB: That sounds like a great idea. They would have the freedom to choose a topic meaningful to them personally. How do you see using a rubric with this idea?

JOAN: The rubric for this exploration project would be a way for us to provide guidance for quality work. Even though the kids will be able to choose their own topics, they should still do their best work in terms of process, product, and presentation.

BARB: I agree, but is there a way we can try something a bit different with the rubric this time? I remember hearing a teacher talk about rubrics she used with her students, and how she eventually got the kids to help her write the criteria. She felt if they helped write the criteria, they would feel more ownership, and therefore take more responsibility for their own work.

JOAN: *(always game for anything)*: Good idea!

NARRATOR: ***Joan and Barb decided to give the students the opportunity to come up with rubric criteria. Handing out blank rubric sheets with only the basic headings, they asked the kids each to fill in the criteria and then discuss their ideas in small groups to come to consensus. Following this, each small group's ideas were***

written on the board in an effort to arrive at agreement on each criteria. This was a rather time-consuming and tedious process but extremely important in giving the students a feeling of ownership.

Joan and Barb, surprised at the similarity of both classes' rubrics, observed that the students seemed to have internalized the criteria from their previous experience. That evening, Joan and Barb combined the criteria statements of the two classes in order to produce a final rubric that would serve as a guide for everyone.

During much of the spring semester, the exploration project time was utilized fairly well by most of the students. For some kids this was their favorite part of the day, and Joan and Barb were awestruck by the high quality and professionalism of many of the performances. Some students took such command of their topics that they took full control of the class for about a half hour while they did their presentation. One group of three girls delved deeply into an exploration of the Holocaust. They read many books and had diligent discussions about the unbelievable occurrences of that time in history. They decided, with the help of an older sister, to write and perform a play.

Other kids dug just as deeply into their areas of interest. One boy did a presentation on medieval times, and another involved his family in helping him made a videotape about ice climbing in which he was the main actor. A boy who was an incredible artist but who had a hard time doing "school things" did a presentation on Michelangelo and Leonardo da Vinci and showed drawings he had done using their works as models. A girl did a presentation about Pearl Harbor that was so professional that she presented it two more times for other classes in the school.

Barb's Classroom

(Barb's classroom during exploration project time. One of several students is having a hard time finding a meaningful topic he can stick with.)

BARB: Tell me some things you're interested in.

STUDENT *(with frustration)*: I don't know.

BARB: Well, tell me what you do when you go home from school.

STUDENT: I ice-skate.

BARB *(with interest)*: Why do you ice-skate?

STUDENT: *(smiling)*: I play hockey.

NARRATOR: ***And there it was, a driving interest. This student consequently brought in a huge bag containing all of his hockey gear to show to the class for his presentation. The students (and Barb) were mesmerized. This rather shy student showed us all his gear, used the blackboard to draw a very clear hockey field, showed us what each player was supposed to do, and answered all our questions about hockey.***

Not all students were so successful, and some went back to their old habits of avoidance. Joan and Barb realized that rubrics weren't working to motivate everyone, so they had the kids keep track of what they got accomplished each day by filling out a calendar. This helped some kids see that they weren't meeting expectations. Some improved; others didn't.

Joan soon came up with a brilliant plan. She asked the kids to look at the rubric they had helped create and to use the criteria to evaluate their peers after each presentation. She modeled this procedure by evaluating one of the presentations aloud while consulting the rubric in her hand.

JOAN: Jimmy, I noticed that you accurately answered our questions. How do you think you could have more actively involved the audience?

NARRATOR: ***Joan also asked the presenters how they felt about the behavior of their audience. Then Joan asked the audience to give the presenter positive feedback followed by suggestions for improvement. The kids gave excellent feedback to each other using this praise/suggestion format, and they often used phrases from the rubric criteria to clarify their suggestions.***

Each child was also required to complete a self-evaluation using the same rubric criteria. Joan and Barb felt that this evaluative side of the rubric was a powerful incentive for students to try to do high-quality work. The more the rubric criteria were used to evaluate their work, the more the kids seemed to rise to the challenge of meeting or going beyond expectations. The students knew they would evaluate themselves as well as be evaluated by their teacher and their peers, and they knew exactly what the standards for a high-quality project entailed.

Joan and Barb were, for the most part, quite pleased with their students' responses to the rubric that they had helped to create. There are, however, always the unexpected responses, and things weren't perfect. Some of the students never bought into the opportunity to do their own exploration. The exploration projects had, for the vast majority of the students, provided an opportunity to further explore their own interests and share them with their peers. Classes were enriched by student presentations on topics they wouldn't ordinarily have had the opportunity to learn about. In addition, the students become more capable of evaluating themselves, their audience, and their peers. The rubric had helped the teachers communicate criteria for higher quality work. Joan and Barb were aware that this was a process, and that processes are often messy. They were learning and so were the kids. Other rubrics were used throughout the semester for a variety of activities and projects in all subject areas.

The Finale ... By Golly!

NARRATOR: ***As a final attempt to understand the role rubrics had played in their classroom, Joan and Barb decided to send a questionnaire home to the parents. They got only a handful back, but they learned that the rubrics had helped the parents understand the expectations for projects more clearly. Some of the parent comments were:***

- ✓ [The rubric] provided a foundation on where to begin and where we were going.
- ✓ I appreciate having the standards in writing because they are clear and available for reference.
- ✓ [Our child] appears to understand what is expected from her and strives to meet, if not exceed, the standards set.
- ✓ [Rubrics] helped us to know what was expected, making it quite clear. [Rubrics] allow for a lot of creative space.

- ✓ We have a much clearer understanding of what you expect.

NARRATOR: ***Joan and Barb decided that one more step was necessary. Wondering what the students thought about the rubrics, they decided to ask. This is what the kids said:***

- ✓ Rubrics are fun...
- ✓ Rubrics are great! They clearly show what the expectations are. They also helped me a lot with keeping organized.
- ✓ Rubrics are like a guideline.
- ✓ The reason I like the rubrics [was] because you know what you are supposed to do, and you know how well you did.
- ✓ I like how [the rubrics] got me to work harder on projects.
- ✓ I like them because I could see how you thought about the project that I did.
- ✓ I do think it helped me evaluate myself and others.
- ✓ I think that to make the rubrics better, the teachers could have let us put in more. They let us make one once, but I think that one was the best because that way, we could say what we thought and it wasn't like a grade, more like comments.
- ✓ Rubrics are helpful because then you know what the teachers expect from you.
- ✓ I liked it because it showed me what to work on and what not to work on.
- ✓ I liked the rubrics because they helped me learn what was expected of me so I would not just rush.
- ✓ I really enjoy the rubric because people that use them get really good feedback on whatever they did, either a project or something else they did. It has helped me because I get input on things that I have maybe skipped or overlooked when I have done my project.
- ✓ Rubrics helped me so I didn't slop together my projects.
- ✓ ... rubrics gave me a goal so I worked harder.

NARRATOR: ***Not all the students' comments were favorable. One child, who incidentally hated to write, "hired" a friend to help him say, "I don't like it. I don't like it because you don't have to put it all down on paper, you can just say it. It doesn't help me. It is a waste of paper and you just forget about it."***

Another student said, "A lot of the time the rubrics didn't do much for me except make me feel bad about certain areas... I don't think you should put sentences on there like I had 'poor spelling, punctuation, grammar, etc.' and you should put things like that more gently."

Someone else said, "The rubrics didn't help me at all." It's interesting, however, that we got the opposite view from this child's mother!
One student said, "With rubrics it's easy to compare with peers and say, 'Mine is better than yours, etc.' Also too many rubrics at once are overwhelming."
Joan and Barb realized that they had only begun their adventure with rubrics. They discovered that the more they knew, the less they knew, but the journey had begun...

For pre-kindergarten adventure, go to Scene II-A (p. 48)
For sixth-seventh grade adventure, go to Scene II-C (below)
For the conclusion, go to The Rest of the Story (p. 60)

SCENE II-C

UNSOLVED MYSTERIES IN THE MIDDLE SCHOOL

(Mid-April. After school in Barb's classroom for a brief meeting of teacher-researchers.)

JANE *(plopping herself down on the couch in Barb's reading area)*: I am so frustrated. I don't think I'll ever get this rubric off the ground. I feel like giving up and going back to just giving the students my objectives for this project. It would be a lot easier.

BARB: What's the matter?

JANE: I can't believe this happened. I was feeling so prepared because of all the work I've seen you and Joanie do with rubrics and all the discussions we've had about how to develop rubrics with the students. I was all psyched to have the students create their own rubric for the Planet Fair. We discussed rubrics, I showed them sample rubrics from various grade levels, and then we began brainstorming categories which students felt were important in the evaluation process. They came up with this great list of ideas, and we were already beginning to develop criteria for evaluation.

JOAN: That was one of the best parts for us, too. The class does know what needs to go into a quality project.

JANE: Yeah, they sure do. In fact, I brought the brainstorm list to the sixth-seventh grade team teachers' meeting to show them what we had begun to do. One of the teachers was so impressed with the list that he asked to take it and type it up on the computer to get it organized. I said, "Sure, thanks," and the team continued on with other business.

CLETA *(nodding)*: Great. It sounds like the rest of your team is interested in using the rubric too.

JANE: Oh boy, the interest was so high that the teacher took the categories, added criteria from another source, copied it off and handed it out to the students during his morning class! The students came into language arts this afternoon telling me they already had a rubric for the Planet Fair and produced the copy to prove it when I protested that we hadn't completed a rubric yet.

JOAN: What? All this was done without the rest of the sixth-seventh team's input?

JANE: That's right, and there goes the student-made rubric for this unit. I just assumed everyone on the team was fine with student-created rubrics, but I assumed wrong. What a bummer. When I asked why the teacher had decided to make the rubric, I was told that, in his opinion, the students were not capable of formulating valid criteria for assessment.

NARRATOR: ***At this point, Audrey, one of the co-facilitators of the teacher/researcher project, stopped by our meeting and sensed the umbrella of gloom hanging over Jane's head. After hearing a brief explanation, Audrey offered her perspective.***

AUDREY: This is an important component of our research. We need to continue to address the issue of collaboration and communication with all people involved in or

affected by our research question. Just include this incident as part of your research and go on. This is valuable stuff.

NARRATOR: ***Wow! A professor who values failure as part of the learning experience? Outstanding. Jane took Audrey's advice: she went on to explain more fully to the sixth-seventh grade team the research question and figured out how to incorporate a student-made rubric into the next unit.***

The Rest of the Story

(Early May. Sixth-seventh grade language arts class. Samples of rubrics from various grade levels and subject areas are laid out on each table as students enter the room.)

NARRATOR: ***The sixth-seventh grade team used the student/teacher-made rubric to plan and assess the Planet Fair projects. Students, parents, and teachers were pleased with the outcome, but the rubric often didn't reflect everything the students had accomplished, or it was repetitive, or too wordy for the students to understand. In other words, this experience became a building block for the development of a completely student-driven rubric. The opportunity came with the next integrated unit, The Ocean. In Jane's language arts class, students had been reading the novel, The Pearl, and were gearing up to write an analysis of one of the characters in Steinbeck's novel.***

ALICE: What are these for, Mrs. Wade? I thought we already did rubrics this year.

JANE: Let's take a look at these samples one more time. Your cooperative group is in charge of making your own rubric which describes the criteria for evaluating this essay on character analysis. You may use your own ideas or borrow ideas from the samples. After each cooperative group develops its own rubric, we will try to come up with a consensus about what this essay needs to look like.

EVAN: You mean we get to decide what our paper should have in it... totally?

JANE: You sure do, Evan. Now, here's the deal. Each of you needs to follow the assignment: an analysis of one of the dynamic characters in the novel. But we are working together to create a plan for evaluating the quality of your essay in the categories we decide are important to assess. Other questions?

MICHELLE: Do we have to use these categories in the samples? What if we want more than just "Exceeds Standard," "Meets Standard," and "Does Not Meet Standard" for our scale?

JANE: This is our rubric. Come up with a plan in your group, and we'll discuss it together.

NARRATOR: ***What was obvious as the groups began working was these middle schoolers knew what a quality essay needed to look like. In fact, they often insisted on tougher criteria than Mrs. Wade would have, particularly in the area of mechanics. Some strong opinions surfaced about the definition of high-quality work, as evidenced in the interchange below:***

STUDENT #1: Why does this rubric say you can only exceed the standard with zero spelling errors? Do you have to be perfect to write a high-quality paper?

STUDENT #2: Hey, just use the spell checker.

STUDENT #3: The doesn't always work. I used the spell checker and still had tons

of errors on my young authors' story when Mrs. Wade edited.

STUDENT #4: Well, you need more than one editor before your final draft.

STUDENT #5: What about if you have trouble with spelling, if it's hard for you?

STUDENT #6: Yeah, and what about grammar? That's the hard part for me. I can spell, but I hate all the grammar rules.

NARRATOR: ***During this interchange, Mrs. Wade stayed in the role of facilitator, encouraging students to come to their own conclusions about what was important in the evaluation of the essay. As the class continued to discuss various categories, they came to consensus on some items and had to resort to majority vote on others. Mrs. Wade noticed how confident each student became in this process. Even students who were not usually strong discussion participants had their group's rubric in front of them so that they could refer to it and emphasize important points for the entire class. Some students were initially confused by the teacher's refusal to just step in and make a decision, but they soon realized everyone was making the rubric together. Students used their power very responsibly; no one tried to get away with poor-quality criteria.***

The rubric development process took three class periods, which included time to assess each other's rubrics and come to a consensus on a class rubric. The rubric aided in the editing and revision process because its criteria became a guide for student decisions about the amount of effort they wanted to put into the assignment. They were not writing to please the teacher but were choosing to meet a specific criteria in content and mechanics.

The sixth-seventh grade team will use this rubric-making process as a springboard next year because it is such an open-ended, ongoing tool for authentic assessment. What is certain is that the rubric will change, it will grow, it will become more valuable as a learning tool. What is still unknown is exactly what the rubric will look like because this form of feedback evolves along with the students, teachers, and parents as all of us continue to work for positive change in the classroom.

ACT III

PUTTING IT ALL TOGETHER

(A warm July day. The four researchers, clad in shorts and sandals, gather in Barb's classroom.)

JANE: Okay, our bodies and brains have had some much-needed relaxation. We ought to be able to put this puzzle together now. Let's go around the table and summarize our conclusions.

CLETA: I've decided that the whole issue, at least for my age group, is not so much about using rubrics as about having clear expectations and high standards and giving specific feedback.

BARB: I agree! Before we created and introduced these rubrics to our students, we hadn't been describing clear parameters of a high-quality project to our stu-

dents. Many of the things our students are now doing well with the help of a rubric, they never had thought of before.

JANE: The rubrics have been consciousness-raising for the kids. Through the use of the rubrics the kids knew exactly what was high quality and what was poor quality. They had to decide how they wanted to perform. The specific feedback provided by the rubric has proven to be very helpful in improving most students' performance. I think it is much clearer assessment tool than traditional grades.

JOAN: As we educated the parents in the use of rubrics, we found that the parents began to examine them with their children and talk together about expectations. The rubrics allowed the kids to set their own goals, putting the responsibility for learning more on the students' shoulders.

NARRATOR: ***Thus, through this sharing dialogue, the team arrived at their first conclusion: Rubrics served as a tool to elevate quality by making expectations clear concerning a student's work. All agreed with the importance of having specific standards, high expectations, and systematic feedback for students.***
BARB: Involving the students in creating the rubric and requiring them to use it for both self-assessment and peer-assessment completes the cycle.

CLETA: What do you mean?

BARB: We used to do the planning and the evaluating. Then we began to include students in the planning. This project has extended that. Now the students are involved in assessment... and in planning the criteria for assessment. Having their opinion of what should be the parameters of a high-quality project taken seriously has given most students power and ownership, and an increased enthusiasm for learning.

CLETA: Yes! I see! It makes assessment an episode of learning, not something separate.

JOAN: The students felt a lot of power when giving feedback to others in the class, and they didn't abuse this power. Most students giving the presentations were much more prepared because they knew they would get honest specific feedback from their audience of peers as well as from the teacher, student teachers, and practicum students.

JANE: When students are required to evaluate themselves, as well as their peers, it puts everyone on a more equal footing, and makes the work itself more important. Teachers as evaluators are no longer put on a pedestal.

NARRATOR: ***Once again, the team had arrived at a very valuable conclusion: Involving the children in the creation of rubrics completed a cycle. Students were now involved in both assessment and in planning the criteria for assessment. Having their input valued gave most students increased ownership and enthusiasm for learning.***

JANE: I've been thinking about the Grant Wiggins (1993a) chapter we read. He says that American teachers tend to avoid telling student how they are performing in relation to a standard. We don't want to damage anyone's self-esteem. We tend to lower our expectations for students and accept their first efforts, which are often much less than their best. Our experience with rubrics has helped us raise our own standards.

CLETA: That's an interesting point. We do need to be positive and encouraging, but the "pat on the back, good job" trend encourages mediocrity.

JOAN: Yes, I think so. Setting high expectations and providing the students with honest feedback has resulted in much higher quality work.

NARRATOR: ***The team had arrived at their third conclusion: Rubrics are one tool for refocusing students, parents, and teachers on a truer understanding of what is useful feedback. Rubrics remind teachers not only to be encouraging, but also to insist on high standards and provide specific feedback about ways student can improve. Rubrics create an increased awareness of the continuous cycle of goal-setting, working, feedback, revision, and further goal-setting that sustains learning.***

CLETA: The main finding from my research is that children, parents, and teachers all value different things about a project. Young children prefer to learn through fantasy, role play, construction, and artistic creation.

BARB: From informal interviews with students about the rubrics, I've also found that kids often value different things than adults. For example, my students think that presentations that are interesting, but short and to the point, are of high quality, as are projects actively involving the audience with things to touch, taste, or fantasize. Our teacherly way of seeing things isn't necessarily the way kids see things.

JANE: Interestingly enough, the middle school students value the same types of learning experiences, and they rate student projects of that type higher on the rubric.

NARRATOR: ***The fourth conclusion was obvious: Paying attention to what children value is important across grade levels.***

JOAN: Rubrics can allow for a lot of different approaches to learning. I've worked a long time with children who have special needs, and I've found that rubrics worked well for those kids in our class here.

JANE: Rubrics can set high standards and still allow many ways for kids to do high-quality work. That's necessary if rubrics are to be useful and fair. I don't want rubrics to be interpreted as requiring every student to do exactly the same thing.

JOAN: There are choices that kids can make to use their strengths. High standards, with specific parameters, as well as a lot of choice in how to reach those standards is important.

BARB: A fourth-fifth grade math experience illustrates those points well. We didn't exactly do rubrics, but we set a standard that each child would show 100% mastery of the multiplication tables by working problem-sets perfectly within a comfortable yet specified time limit. They could keep working on this task until they had mastered it. I had one child in particular who was overwhelmed by having to write down the answers to a whole page of multiplication facts on a timed test. Sensing his frustration, I allowed him to complete the page one line at a time, orally. When his total time was added up, he was one of the fastest in the class.

JOAN: I had another student who had a severe case of text anxiety. When I took the stopwatch away and told her not to worry about the time, she completed the task in much less than the specified two minutes. She was unaware that I had glanced at the second hand on the wall clock just to check.

JANE: Yes, having different ways to meet the standards is vital for all kids, not just those with special needs. Allowing diversity in the ways students arrive at objectives allows students to demonstrate knowledge using many different strengths or intelligences.

NARRATOR: ***The team had, through their discussions, come up with another important conclusion: To be useful and fair, rubrics have to set specific standards***

but allow room for variations in the ways students meet those standards.

CLETA: I'd like to change the subject a bit. Even though we were all doing different things, and mine was the most different, I found it really helpful to have the three of you, individually and as a group, to talk things over with. You really helped me figure out what I was doing. I think that is my most important conclusion from all this research.

JANE: Mine too. It didn't matter that our projects weren't the same. We could still get helpful feedback from each other.

JOAN: I really want to keep up our discussion. Wouldn't it be great if we could get more of the faculty to participate?

BARB: This project has encouraged me to read more, and it's been helpful to get everyone's perspective on what I'm reading. I'd like to start a faculty book-discussion group to meet regularly. We're starting to use Howard Gardner's idea of multiple intelligences as a framework for assessment and reporting throughout the school. We need to all be reading and talking about this aspect of our school philosophy.

JANE: Yes, and each one of us has special areas of knowledge. Just think what we'd learn if each person took a turn suggesting something we should all read, then leading the discussion on that subject. It wouldn't even have to be long books; an article or excerpt would do to get us started. The shared thinking and talking is what's important!

CLETA: I'd also like to see us continue some shared research projects. I've learned so much from the team's encouragement to experiment and expand my thinking.

BARB: Okay, we've all agreed. We'll find some way to continue discussion and collaboration, maybe by reading and discussing ideas as we try them in our classroom. And... we'll open this regular collaborative group up to other faculty who want to join. In fact, we'll encourage them to participate.

NARRATOR: ***The teacher-research team had just reached what they considered their most important conclusion: It is essential for teachers to meet regularly and talk together about common efforts to make positive changes in their classrooms. Teachers learn a great deal from each other.***

EPILOGUE

LOOKING AHEAD TO THE SEQUEL: SCARY STORIES FROM THE CLASSROOM, PART II

CLETA: Do you think there's any danger that rubrics could turn into the old letter-grade system in disguise?

JOAN: Good point! Maybe rubrics shouldn't have a below, meets, or exceeds expectations component that is so similar to letter grades.

JANE: With rubrics designed that way, some students are still content to do the least they can to get by. They don't care if the assessments tell them they did work that was "below expectations." I'm not sure what to do to motivate these students.

BARB: Maybe the rubric should only give very clear descriptions of different aspects

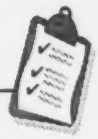
of excellent work, and we should insist that every student meet these expectations.

JOAN: And if the expectations weren't met, the student would be required to revise, edit, or redo until a high-quality product was achieved. In this way both the process and the product are of equal importance, and it would be much more likely that all students would be motivated to produce high-quality work.

CLETA: What about the issue of parents, kids, and teachers having their own agendas for what makes a good project? How do we take parents' views into account? And how can the use of more imagination and arts at all age levels help children learn?

NARRATOR: ***This team of teachers has clearly only scratched the surface of classroom research. It seems that the more they discover about themselves, their students, and their classrooms, the more obvious it becomes that they have much more to learn through their collaboration.***

Chapter 7



Creating Rubrics Through Negotiable Contracting

Andi Stix
The Interactive Classroom, New Rochelle, New York

Chapter 6 put the spotlight on teachers at three different grade levels as they explored scoring rubrics as part of their action research projects. The teachers concluded that rubrics were valuable for making expectations clear, increasing student ownership, and prompting useful feedback. Rubrics also provided standards while allowing multiple approaches for meeting them. In this chapter, Andi Stix details an approach to involving students in the creation of rubrics that she calls "negotiable contracting." She finds that students are conscientious and rigorous in the design of rubrics, including a rubric for a geography mural and one for an original poem.

What would happen if students were invited to help decide how their work should be evaluated? Would they exploit the opportunity, designing standards ridiculously low to ensure a glut of effortless good grades?

Surprisingly, the answer is no. Experience at Robert Wagner Middle School in Manhattan shows that students who are given a role in the assessment process can and do rise to the occasion. Given the appropriate direction by their teachers, youngsters are able to evaluate their strengths and weaknesses accurately and pinpoint where

---

In addition to owning and operating an educational consulting firm (http//www.interactiveclassroom.com), Professor Stix hold a part-time position as adjunct full professor at Pace University in Pleasantville, New York, and specializes in constructivist education. This chapter is derived from the following sources: *Strategies for Student-Centered Assessment* (1996) and *A Rubric Bank for Teachers* (1999), both available from The Interactive Classroom, as well as *Empowering Students Through Negotiable Contracting*, a paper presented at the 1997 National Middle School Initiative Conference in Long Island, NY (ERIC Document Number ED 411 274) and *Creating Rubrics Through Negotiable Contracting and Assessment* (with Michele Block Morse), a paper presented at the 1996 National Middle School Conference in Baltimore, MD (ERIC Document Number ED 411 273).

to focus their efforts to get the most out of what they're learning. As a result, students come to view assessment not as an arbitrary form of reward or humiliation (a common perception of middle school students), but as a positive tool for personal growth.

Negotiable contracting, an approach to involving students in the assessment process, is being implemented in some schools in the New York City area. Negotiable contracting is adaptable to both arts and science curricula and is flexible enough to accommodate multi-modal forms of learning.

Empowering Students

The art of negotiable contracting consists of giving students shared ownership in their own learning (Wiggins, 1993b). Although the teacher is ultimately responsible for grading, he or she functions not as an all-powerful judge of students' work, but as a facilitator of discussion on the assessment process (Seeley, 1994). With negotiable contracting, before teachers present their own expectations of the work, they ask students to give their opinions about what would constitute high-quality work. It can be helpful to show students examples of the work to help them formulate their ideas about quality. Across the "negotiating table," teachers and their classes arrive at a consensus that is mutually acceptable. Because students feel themselves to be valued participants in the assessment process, they are motivated to strive toward the criteria-based standards.

The contract process can be used independently of a formal evaluation and can serve a variety of purposes. For example, if students are to work together in groups, negotiable contracting is helpful in setting up expectations regarding cooperative roles, research materials, and presentation formats.

Creating the Rubric

The rubric is an important element of using negotiable contracting for formal assessment (Pate, Homestead, and McGinnis, 1993). A rubric is a carefully designed ratings chart drawn up jointly by teachers and students. Unlike a traditionally assigned, generalized number or letter grade, the rubric serves as an in-depth "report card" for a lesson, unit, or project. Along one side of the rubric are listed the criteria that the teachers and students decide are the most important to be mastered in the lesson. Three to five criteria for each task are generally manageable. Across the top of the rubric are listed the rankings that will be used to assess how well students understand or demonstrate each of the criterion. Choosing neutral words for each rating avoids the implication of good/bad inherent in a traditional A through F or numerical grading system. The state of Kentucky, which uses a rubric system of assessment, has chosen these non-pejorative ratings: Novice, Apprentice, Proficient, and Distinguished. Working with an even number of ratings helps teachers and students avoid the natural temptation of awarding the middle grade, such as 3 on a 1 to 5 ranking system. Within each ranking, there may also be numerical gradations, depending on whether a student performs on the higher or lower level of that category. Inside the rubric, the criteria for each level of achievement are explained in detail. (Interested readers may wish to contact the Interactive Classroom at www.interactiveclassroom.com to find *A Rubric Bank for Teachers*, which lists 100 highly detailed criteria for rubric assessment.)

A Social Studies Example

Let's take as an example a social studies teacher, Ms. Polin, who assigned her students at Robert Wagner Middle School the task of creating a mural for a geography lesson. Before they began any work on the murals, she arranged the class in cooperative learning groups and asked them to consider the question, "If you were me, what qualities would you look for in deciding how to grade each mural? Come up with six criteria that you would look for." After allowing time for discussion, Ms. Polin asked each group to rank the qualities they had selected in order of importance, from most important to least important.

Next, each group presented its top two criteria to the class. Ms. Polin listed those criteria on the board and the class was asked to choose which ones were truly most relevant to the lesson. With the teacher's guidance, they agreed on three qualities: 1) detail and depth; 2) a clear focal point; and 3) high-quality design. They were then asked what should be considered poor, fair, good, and excellent performance for each criterion. One student suggested that a poor mural would have most of the facts wrong, and the other students readily agreed. "What about if only some of the facts are wrong?" Ms. Polin asked. "That would be a fair grade," said one boy. "I think having some of the facts wrong should still be a poor grade," argued another student. Finally, after more discussion, a consensus was reached among the class that getting only some of the facts wrong would earn a 'fair' grade. After more discussion, they also decided that getting all the facts right should earn a 'good' rating, while gathering an exceptional amount of accurate, interesting information from unusual sources would earn a rating of 'excellent.'

As a result of their negotiations, before they even picked up a pencil or pen, Ms. Polin's students were perfectly clear about what was expected in their murals. Moreover, they had the satisfaction of having had a voice in setting the objectives for the project and establishing a rating system that they considered to be fair. Figure 1 shows the rubric they created.

Figure 1: Rubric for a Geography Mural

	Novice	Apprentice	Veteran	Master
Detail and Depth	Incorrect or few facts; hardly any detail (1 to 3 points)	Some facts are accurate; some detail is included (4 to 6 points)	Substantial number of facts; good amount of detail (7 to 9 points)	Exceptional number of facts; vivid descriptions (10 to 12 points)
Focus	Vague and unclear (1 to 2 points)	Some focus, but not organized enough (3 to 4 points)	Well organized and clearly presented (5 to 6 points)	Highly organized and easy to follow (7 to 8 points)
Design	Little or no layout and design (1 to 3 points)	Simple design; layout could be more organized (4 to 6 points)	Attractive and inviting to the viewer (7 to 9 points)	Exceptional design and outstanding visual appeal (10 to 12 points)

BEST COPY AVAILABLE

Note that there is no overall rating for the child; the terms are used separately to evaluate students' performance on each of the criterion in the rubric. If students work in groups, a rubric might also be used to address behavioral aspects of group problem solving, such as listening, taking turns, and sharing materials.

To save time during the rubric development process, a spokesperson from each group might bring forward a single criterion as group members check off items on their list that have already been mentioned to eliminate repetition. It might happen that the teacher has a criterion that was not posted on the board but is essential to a fair and equitable assessment of the project. If that occurs, the teacher can explain to the class that an additional item is being added to the list and provide details about why the addition is meaningful from the instructor's point of view. After all the groups have submitted their ideas, the students can discuss them, then return to their cooperative groups to prioritize their top five. The class can combine their lists, noting which criteria are heavily weighted by virtue of appearing on many groups' lists, and which are not. It is recommended that four or five criteria be selected to formulate a rubric. More than that number might be overwhelming to students and are not likely to be necessary.

More Examples

Rubrics can be used for any subject or lesson that requires students to demonstrate their competence. They can be applied to journal work, projects, research studies, experiments, and skits. Rubrics can be especially effective in assessing students' work in mathematics (Moon, 1993). While rote skills such as memorizing the multiplication tables may be best suited to traditional quizzing and grading, the majority of mathematics involves creative problem solving in which there are several ways to arrive at a solution, some more succinct, effective, or creative than others. For a lesson involving word problems in fractions, for example, the "report card" for students' problem solving might include the following assessment criteria decided upon by teacher and students: Is the solution easy to follow? Does it demonstrate clear conceptual understanding? Would the answer work in real life? Do the diagrams, sentences, and numbers coordinate?

Let's look at how a rubric would be utilized in Mrs. Bartko's eighth-grade language arts class, which is studying a unit on poetry. After discussing how poetry differs from prose and looking at various types of poetry, the students are assigned to write a poem of their own. Mrs. Bartko then asks, "How can a poem—a subjective assignment with no 'correct' answer-be assessed fairly?"

The students launch into a discussion of what constitutes good poetry. Working in groups, they come up with a rubric, composed of four main criteria that Mrs. Bartko and the students agree are the most appropriate and fair qualities. They decide a poem should portray emotion and/or imagery, captivate the reader, use language clearly, and use punctuation purposefully. Mrs. Bartko and her students then read various examples to see how those skills are applied at the various ratings levels. The rubric they devise is shown in Figure 2.

Figure 2: Rubric for an Original Poem

	Novice	Apprentice	Veteran	Master
Ability to Captivate the Reader	Unfocused: author seems unsure of direction (1 to 2 points)	Some focus, but lacks continuity (3 to 4 points)	Well focused and interests reader throughout (5 to 6 points)	Captivates and involves reader deeply (7 to 8 points)
Sensory Images	Difficult to visualize image or emotion (1 to 3 points)	Some use of image, idea, or emotion (4 to 6 points)	Clear use of sensory images to portray ideas or emotions (7 to 9 points)	Vivid, detailed images and intensely felt emotion (10 to 12 points)
Use of Language	Imprecise or inappropriate choice of words (1 to 2 points)	Expresses thoughts marginally (3 to 4 points)	Appropriate choice of language (5 to 6 points)	Uses rich and imaginative language (7 to 8 points)

In addition to the rubric itself, there is an area included for comments. In this space, Mrs. Bartko can be even more specific about the strengths and weaknesses of individual student's work and suggest ways to stretch skills and expand understanding.

At Robert Wagner Middle School, some teachers have enlarged a blank rubric and laminated it. For each project, they use a dry-erase marker and fill in the quadrants with the students. The students can use a blank sheet to create their own record of what is expected of them. At the end of the project, they may be asked to assess themselves or their peers using the assessment sheet.

End-of-Year Assessment

Along with the rubrics developed for individual assignments and lessons, each student's assessment should encompass an overall look at how far the student has come during the year and what his or her strengths and weaknesses are. Throughout the year, the teacher might periodically ask students to select lessons or assignments that they have found particularly significant and explain why in writing. Students might choose examples that functioned as benchmarks in their understanding or that were particularly interesting or challenging. As an alternative, the teacher might ask students to choose three examples of their best work, and one considered substandard, and explain the rationale behind their choices. The examples of student work could then be collected into a portfolio for an end-of-year assessment. The portfolio demonstrates vividly for teachers, students, and parents alike how the youngster's thinking has evolved over the course of the year.

Conclusion

Students who are involved in developing and applying a rubric become clear about what skills they need to master and how well they are progressing. They develop confidence in their abilities and the incentive to push on when they run into difficulties. Rubrics also help teachers assess and grade students with fairness, honesty, and complete understanding of the qualities that constitute excellent work.

Chapter 8

Designing Scoring Rubrics for Your Classroom

Craig A. Mertler
Bowling Green (OH) State University

The previous chapter describes how students in a New York middle school were involved in the creation of rubrics by which their work would be assessed. If teachers believe that important criteria are being left out, they can introduce them as part of the process of negotiable contracting. Teachers who are new to rubrics may wish to do most of the development by themselves, at least at first, by drawing on their knowledge of their subjects and any relevant state standards. The step-by-step guide to designing scoring rubrics presented in this chapter seems a fitting close to the book because it will help you review such issues as choosing between holistic and analytic rubrics and converting rubric scores to grades and will help you begin putting your knowledge into practice. Be open to reevaluating and revising your scoring rubrics in order to improve their quality and effectiveness.

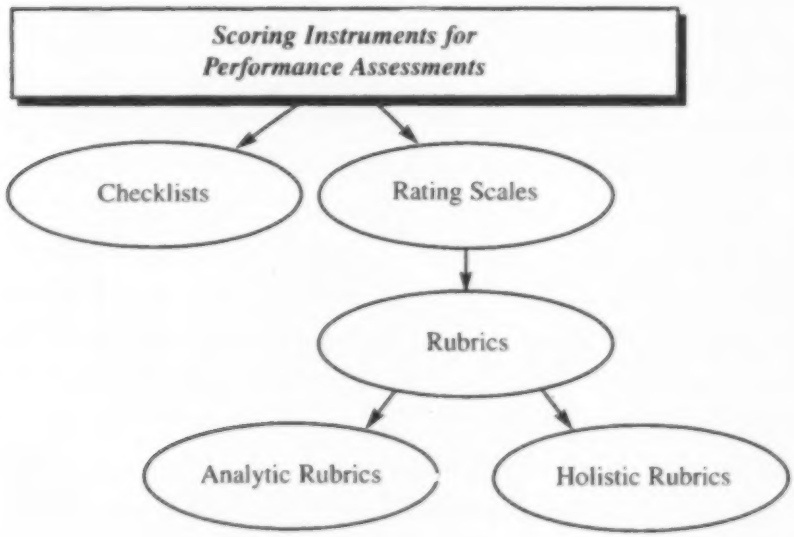
Rubrics are rating scales—as opposed to checklists—that are used with performance assessments. They are formally defined as scoring guides, consisting of specific pre-established performance criteria, used in evaluating student work on performance assessments. Rubrics are typically the specific form of scoring instrument used when evaluating student performances or products resulting from a performance task.

There are two types of rubrics: holistic and analytic (see Figure 1). A holistic rubric requires the teacher to score the overall process or product as a whole, without judging the component parts separately (Nitko, 2001). In contrast, with an

A slightly longer version of this chapter first appeared in the online, peer-reviewed journal, *Practical Assessment, Research & Evaluation*, 7 (25), available at http://ericae.net/pare/.

analytic rubric, the teacher scores separate, individual parts of the product or performance first, then sums the individual scores to obtain a total score (Moskal, 2000; Nitko, 2001).

Figure 1: Types of Scoring Instruments for *Performance Assessments*



Holistic rubrics are customarily utilized when errors in some part of the process can be tolerated provided the overall quality is high (Chase, 1999). Nitko (2001) further states that use of holistic rubrics is probably more appropriate when performance tasks require students to create some sort of response and where there is no definitive correct answer. The focus of a score reported using a holistic rubric is on the overall quality, proficiency, or understanding of the specific content and skills—it involves assessment on a unidimensional level (Mertler, 2001). Use of holistic rubrics can result in a somewhat quicker scoring process than use of analytic rubrics (Nitko, 2001). This is basically due to the fact that the teacher is required to read through or otherwise examine the student product or performance only once, in order to get an "overall" sense of what the student was able to accomplish (Mertler, 2001). Since assessment of the overall performance is the key, holistic rubrics are also typically, though not exclusively, used when the purpose of the performance assessment is summative in nature. At most, only limited feedback is provided to the student as a result of scoring performance tasks in this manner. A template for holistic scoring rubrics is presented in Table 1.

Table 1: Template for Holistic Rubrics

Score	Description
5	Demonstrates complete understanding of the problem. All requirements of task are included in response.
4	Demonstrates considerable understanding of the problem. All requirements of task are included.
3	Demonstrates partial understanding of the problem. Most requirements of task are included.
2	Demonstrates little understanding of the problem. Many requirements of task are missing.
1	Demonstrates no understanding of the problem.
0	No response/task not attempted.

Analytic rubrics are usually preferred when a fairly focused type of response is required (Nitko, 2001); that is, for performance tasks in which there may be one or two acceptable responses and creativity is not an essential feature of the students' responses. Furthermore, because analytic rubrics result initially in several scores, followed by a summed total score, their use represents assessment on a multidimensional level (Mertler, 2001). Both the construction and use of analytic rubrics can be quite time-consuming. A general rule of thumb is that an individual's work should be examined a separate time for each of the specific performance tasks or scoring criteria (Mertler, 2001). However, the advantage to the use of analytic rubrics is quite substantial. Students (and teachers) receive specific feedback on their performance with respect to each of the individual scoring criteria—something that does not happen when using holistic rubrics (Nitko, 2001). It is possible to then create a "profile" of specific student strengths and weaknesses (Mertler, 2001). A template for analytic scoring rubrics is presented in Table 2.

Table 2: Template for Analytic Rubrics

	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
Criteria #1	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #2	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #3	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #4	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	

Prior to designing a specific rubric, a teacher must decide whether the performance or product will be scored holistically or analytically (Airasian, 2000 and 2001). Regardless of which type of rubric is selected, specific performance criteria and observable indicators must be identified as an initial step to development. The decision regarding the use of a holistic or analytic approach to scoring has several possible implications. The most important of these is that teachers must consider first how they intend to use the results. If an overall, summative score is desired, a holistic scoring approach would be more desirable. In contrast, if formative feedback is the goal, an analytic scoring rubric should be used. It is important to note that one type of rubric is not inherently better than the other—you must find a format that works best for your purposes (Montgomery, 2001). Other implications include the time requirements, the nature of the task itself, and the specific performance criteria being observed.

As the templates in Tables 1 and 2 illustrate, the various levels of student performance can be defined using either quantitative (i.e., numerical) or qualitative (i.e., descriptive) labels. In some instances, teachers might want to utilize both quantitative and qualitative labels. If a rubric contains four levels of proficiency or understanding on a continuum, quantitative labels would typically range from 1 to 4. When using qualitative labels, teachers have much more flexibility and can be more creative. A common type of qualitative scale might include the following labels:

BEST COPY AVAILABLE

master, expert, apprentice, and novice. Nearly any type of qualitative scale will suffice, provided it "fits" with the task.

One potentially frustrating aspect of scoring student work with rubrics is the issue of somehow converting scores to grades. It is not a good idea to think of rubrics in terms of percentages (Trice, 2000). For example, if a rubric has six levels (or "points"), a score of 3 should not be equated to 50% (an "F" in most letter grading systems). The process of converting rubric scores to grades or categories is more a process of logic than it is a mathematical one. Trice (2000) suggests that in a rubric scoring system, there are typically more scores at the average and above- average categories (i.e., equating to grades of "C" or better) than there are at the below-average categories. For instance, if a rubric consisted of nine score categories, the equivalent grades and categories might look like this:

Table 3: Sample Grades and Categories

Rubric Score	Grade	Category
8	A+	Excellent
7	A	Excellent
6	B+	Good
5	B	Good
4	C+	Fair
3	C	Fair
2	U	Unsatisfactory
1	U	Unsatisfactory
0	U	Unsatisfactory

When converting rubric scores to grades (typical at the secondary level) or descriptive feedback (typical at the elementary level), it is important to remember that there is not necessarily one correct way to accomplish this. The bottom line for classroom teachers is that they must find a system of conversion that works for them and fits comfortably into their individual system of reporting student performance.

Steps in the Design of Scoring Rubrics

A step-by-step process for designing scoring rubrics for classroom use is presented below. Information for these procedures was compiled from various sources (Airasian, 2000 and 2001; Mertler, 2001; Montgomery, 2001; Nitko, 2001; Tombari and Borich, 1999). The steps will be summarized and discussed, followed by presentations of two sample scoring rubrics.

Step 1: *Re-examine the learning objectives to be addressed by the task.* This allows you to match your scoring guide with your objectives and actual instruction.

Step 2: *Identify specific observable attributes that you want to see (as well as those you don't want to see) your students demonstrate in their product, process, or performance.* Specify the characteristics, skills, or behaviors that you will be looking for, as well as common mistakes you do not want to see.

Step 3: *Brainstorm characteristics that describe each attribute.* Identify ways to describe above-average, average, and below-average performance for each observable attribute identified in Step 2.

Step 4a: *For holistic rubrics, write thorough narrative descriptions for excellent work and poor work incorporating each attribute into the description.* Describe the highest and lowest levels of performance combining the descriptors for all attributes.

Step 4b: *For analytic rubrics, write thorough narrative descriptions for excellent work and poor work for each individual attribute.* Describe the highest and lowest levels of performance using the descriptors for each attribute separately.

Step 5a: For holistic rubrics, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for the collective attributes. Write descriptions for all intermediate levels of performance.

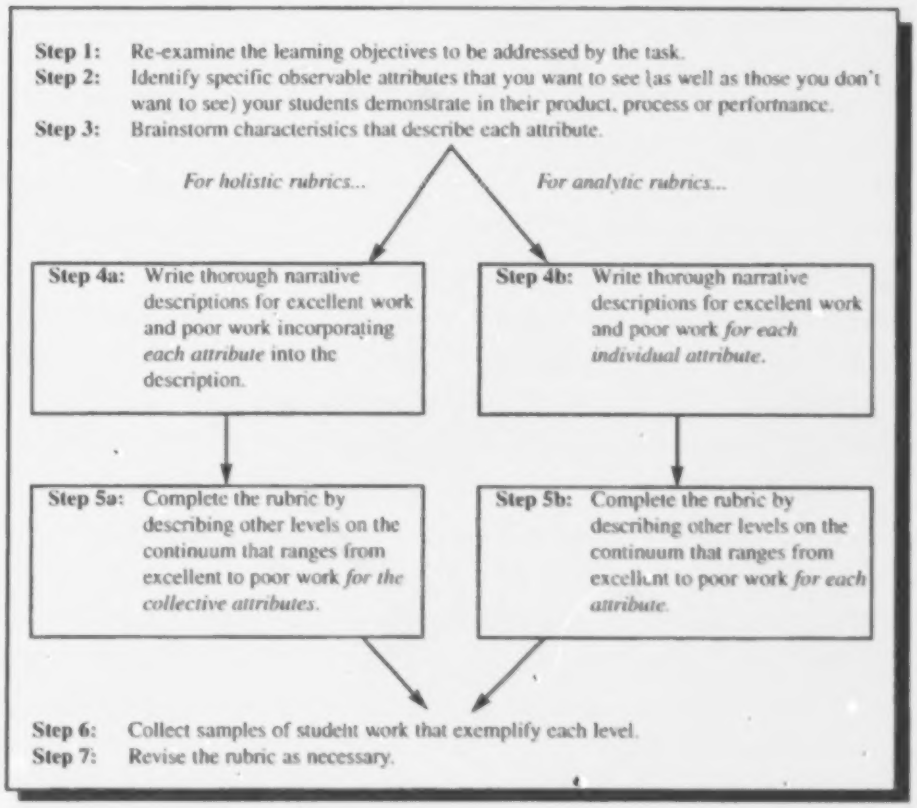
Step 5b: *For analytic rubrics, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for each attribute.* Write descriptions for all intermediate levels of performance for each attribute separately.

Step 6: *Collect samples of student work that exemplify each level.* These will help you score in the future by serving as benchmarks.

Step 7: *Revise the rubric, as necessary.* Be prepared to reflect on the effectiveness of the rubric and revise it prior to its next implementation.

These steps involved in the design of rubrics have been summarized in Figure 2 below.

Figure 2: Designing Scoring Rubrics: Step-By-Step Procedures


Step 1: Re-examine the learning objectives to be addressed by the task.
Step 2: Identify specific observable attributes that you want to see (as well as those you don't want to see) your students demonstrate in their product, process or performance.
Step 3: Brainstorm characteristics that describe each attribute.
For holistic rubrics...
For analytic rubrics...
Step 4a: Write thorough narrative descriptions for excellent work and poor work incorporating each attribute into the description.
Step 4b: Write thorough narrative descriptions for excellent work and poor work for each individual attribute.
Step 5a: Complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for the collective attributes.
Step 5b: Complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for each attribute.
Step 6: Collect samples of student work that exemplify each level.
Step 7: Revise the rubric as necessary.

Two Examples

Two sample scoring rubrics corresponding to specific performance assessment tasks are presented next. Brief discussions precede the actual rubrics. For illustrative purposes, a holistic rubric is presented for the first task and an analytic rubric for the second. It should be noted that either a holistic or an analytic rubric could have been designed for either task.

BEST COPY AVAILABLE

Example 1: Subject - Mathematics
Grade Level(s) - Upper Elementary

Mr. Harris, a fourth-grade teacher, is planning a unit on the topic of data analysis, focusing primarily on the skills of estimation and interpretation of graphs. Specifically, at the end of this unit, he wants to be able to assess his students' mastery of the following instructional objectives:

- ✓ Students will properly interpret a bar graph.
- ✓ Students will accurately estimate values from within a bar graph. (step 1)

Since the purpose of his performance task is summative in nature—the results will be incorporated into the students' grades—he decides to develop a holistic rubric. He identifies the following four attributes on which to focus his rubric: estimation, mathematical computation, conclusions, and communication of explanations (steps 2 and 3). Finally, he begins drafting descriptions of the various levels of performance for the observable attributes (steps 4 and 5). The final rubric for his task appears in Table 4.

Table 4: Math Performance Task - Scoring Rubric Data Analysis

Name _____ Date _____

Score	Description
4	Makes accurate estimations. Uses appropriate mathematical operations with no mistakes. Draws logical conclusions supported by graph. Sound explanations of thinking.
3	Makes good estimations. Uses appropriate mathematical operations with few mistakes. Draws logical conclusions supported by graph. Good explanations of thinking.
2	Attempts estimations, although many inaccurate. Uses inappropriate mathematical operations, but with no mistakes. Draws conclusions not supported by graph. Offers little explanation.
1	Makes inaccurate estimations. Uses inappropriate mathematical operations. Draws no conclusions related to graph. Offers no explanations of thinking.
0	No response/task not attempted.

Example 2: Subjects - Social Studies; Probability and Statistics
Grade Level(s) - 9 - 12

Mrs. Wolfe is a high school American government teacher. She is beginning a unit on the electoral process and knows from past years that her students sometimes have difficulty with the concepts of sampling and election polling. She decides to give her students a performance assessment so they can demonstrate their levels of understanding of these concepts. The main idea that she wants to focus on is that samples (surveys) can accurately predict the viewpoints of an entire population. Specifically, she wants to be able to assess her students on the following instructional objectives:

- ✓ Students will collect data using appropriate methods.
- ✓ Students will accurately analyze and summarize their data.
- ✓ Students will effectively communicate their results. (step 1)

Since the purpose of this performance task is formative in nature, she decides to develop an analytic rubric focusing on the following attributes: sampling technique, data collection, statistical analyses, and communication of results (steps 2 and 3). She drafts descriptions of the various levels of performance for the observable attributes (steps 4 and 5). The final rubric for this task appears in Table 5.

Table 5: Performance Task - Scoring Rubric Population Sampling

Name _____ Date _____

	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
Sampling Technique	Inappropriate sampling technique used.	Appropriate technique used to select sample; major errors in execution.	Appropriate technique used to select sample; minor errors in execution.	Appropriate technique used to select sample; no errors in procedures.	
Survey/ Interview Question	Inappropriate questions asked to gather needed information.	Few pertinent questions asked; data on sample is inadequate.	Most pertinent questions asked; data on sample is adequate.	All pertinent questions asked; data on sample is complete.	
Statistical Analyses	No attempt at summarizing collected data.	Attempts analysis of data, but uses inappropriate procedures.	Proper analytical procedures used, but analysis is incomplete.	All proper analytical procedures used to summarize data.	
Communication of Results	Communication of results is incomplete, unorganized, and difficult to follow.	Communicates some important information; not organized well enough to support decision.	Communicates most of important information; shows support for decision.	Communication of results is very thorough; shows insight into how data predicted outcome.	
				Total Score =	

BEST COPY AVAILABLE

Closing Thoughts

If, as a classroom teacher, you are new to the skill of developing scoring rubrics, do not be discouraged as a result of your initial attempts to design them. As with anything new, there is some degree of trial and error involved. Teachers must be flexible and willing to revisit their self-developed rubrics from time to time. After a rubric is developed and used for the first time, it is not uncommon for teachers to discover gaps in the rubric—that is, aspects of the task that were not initially considered, but ultimately emerged in the students' products. Obviously, it is then imperative for the teacher to revise the rubric accordingly. Furthermore, a rubric designed and used during a given school year may work extremely well for that particular group of students, but not for next year's students. This often occurs as a result of the academic and behavioral characteristics, as well as the group dynamics, of a particular class. In these cases, rubrics must again be revised and adjusted-or perhaps, more comprehensively reworked-in order to appropriately meet the instructional and assessment needs of the particular group of students. This process of re-evaluating and revising self-developed scoring rubrics results in the development of higher quality and more effective scoring rubrics for classroom use. In addition, this process engages the classroom teacher in the instructional process by focusing attention on the alignment of assessment with instructional content and skills.

Scoring Rubrics Resources and Review

Carol Boston
ERIC Clearinghouse on Assessment and Evaluation
University of Maryland

Online Resources (current as of February 2002)

Chicago Public Schools Rubric Bank
http://intranet.cps.k12.il.us/Assessments/Ideas_and_Rubrics/Rubric_Bank/rubric_bank.html
The rubric bank contains PDF files with dozens of analytic and holistic rubrics used in reading, mathematics, science, social studies, the fine arts, speaking, and writing from multiple sources.

CyberLibrary
http://www.rainbowtech.org/CyberLib/assess.htm
This assessment and rubrics portal, maintained by educator Shari Barnhart, offers links to general articles and subject-specific rubrics.

ERIC Clearinghouse on Assessment and Evaluation
http://ericae.net/faqs/rubrics/scoring_rubrics.htm
ERIC/AE offers a response to frequently asked questions about scoring rubrics and their definition and construction. The FAQ includes a brief commentary, links to online resources, a bibliography, and dynamic searches of the ERIC database with descriptions of relevant documents and journal articles.

Kathy Shrock's Guide for Educators
http://school.discovery.com/schrockguide/assess.html

This extensive collection of links will help teachers locate online collections of subject-specific and general rubrics as well as tools to generate their own rubrics. Also included are rubrics for evaluating student Web pages and teacher use of technology.

National Center on Student Standards, Evaluation, and Testing
http://cresst96.cse.ucla.edu/CRESST/pages/Rubrics.htm
The CRESST Web site presents rubrics for evaluating writing mechanics, content knowledge, short answer responses, and problem solving tasks.

Northwest Regional Educational Laboratory
http://www.nwrel.org/assessment/
NWREL's site includes the 6 + 1 Traits Writing rubric developed by the Lab as well as rubrics for reading, Spanish writing, and oral communication. A 700-item online assessment library may be searched for additional materials about scoring rubrics.

Performance Assessment Links in Science (PALS)
http://pals.sri.com/index.html
PALS is an online standards-based resource bank of science performance assessment tasks and scoring rubrics from a range of sources, including the National Assessment of Educational Progress, the Council of Chief State School Officers, and state departments of education in Kentucky, New York, and Oregon. The National Science Foundation sponsors the site.

Prince George's County (MD) Public Schools
http://www.pgcps.pg.k12.md.us/~elc/developingtasks.html
This Web page summarizes the characteristics of performance-based assessments and provides links to other pages that describe the learning theory behind them. It also provides a step-by-step guide to performance-assessment development and demonstrate scoring by using an example from the Maryland School Performance Assessment Program.

Rubistar
http://rubistar.4teachers.org/
This Web site, sponsored by High Plains Regional Technology in Education Consortium, enables teachers to view and customize existing rubrics for project-based learning. Teachers select a performance task, such as making an oral presentation, writing a lab report, or designing a Web site, then select from a list of traits that might be evaluated. The rubric maker will suggest sample descriptions for various levels of performance, which the teacher is free to edit.

Rubric Generators
http://www.teach-nology.com/web_tools/rubrics/
Teachnology, Inc., a New York-based consulting firm, maintains a Web site that includes already-prepared rubrics for various skills and subject areas (e.g., oral presentation, timeline, science fair project) as well as a tool that enables teachers to create their own rubrics. Teachers are encouraged to look carefully at the rubrics to ensure that the proficiency levels and point values described in some of the rubrics match their own priorities and expectations.

Staff Room for Ontario's Teachers
http://www.odyssey.on.ca/~elaine.coxon/rubrics.htm
This Web site offers an extensive collection of rubrics, organized by subject areas and skill areas, including rubrics for group work, visual arts, dance, drama, and foreign language. The site also includes several rubrics for teachers about evaluating lesson plans.

Review

Before you design any assessment, ask yourself these questions:

- ✓ What concept, skill, or knowledge am I trying to assess?
- ✓ What should my students know?
- ✓ At what level should my students be performing?
- ✓ What type of knowledge is being assessed: reasoning, recall, process, or skill?

Scoring rubrics are a valuable assessment method because they

- ✓ Measure progress toward the achievement of complex learning targets such as problem solving and communication;
- ✓ Spell out the criteria to be used for evaluating a product or a task, thus helping students internalize the standards for high quality and teachers make fair and accurate judgments about student achievement and areas for further growth; and
- ✓ Increase student and teacher engagement in learning.

To learn more about the effects of rubrics on student achievement and engagement, see p. 15.

Two types of rubrics are:

- ✓ Analytic rubrics, which yield a score based on student performance on several discrete factors.
- ✓ Holistic rubrics, which yield a single score based on an overall impression of quality.

See p. 8 and pp. 73-75 for suggestions on when to use each type of rubric.

You can use an existing rubric, adapt an existing rubric, or create a completely new rubric. Students can also participate in the creation of rubrics. Be sure to evaluate your rubric on such dimensions as its coverage of important content, the generalizability of its results, the ease with which it can be understood and applied, and its fairness and lack of bias.

For tips on selecting and developing scoring rubrics, see pp. 9-10, as well as the metarubric starting on p. 16.

Learn how students can be involved in creating rubrics through a process called negotiable contracting starting on p. 66.

You'll want to keep in mind such technical considerations as validity and reliability so that rubrics measure what you intend them to measure and results are accurate and stable.

For a brief discussion of technical issues, see p. 12; a more detailed discussion is available on pp. 25-33.

Many methods exist to convert rubric scores to grades if necessary. There is no perfect translation system. Teachers are encouraged to consider various approaches as they develop rubrics.

> *Think through the pros and cons of four methods of assigning grades to rubrics in the case study starting on p. 34. An example grade scale is provided on p. 76.*

Implementing rubrics can be a complex matter. Being willing to experiment, learn from mistakes, and draw on existing resources will help you be successful.

> *Follow a team of teachers as they implement rubrics in their classrooms in the case study presented on p. 41.*
>
> *Get step-by-step design instructions for your own classroom starting on p. 72.*
>
> *Go online to locate more great examples and background materials. Web addresses start on p. 82.*

References

Airasian, P.W. (1991). *Classroom Assessment.* New York: McGraw-Hill.

Airasian, P.W. (2000). *Assessment in the Classroom: A Concise Approach* (2nd ed.). Boston: McGraw-Hill.

Airasian, P. W. (2001). *Classroom Assessment: Concepts and Applications* (4th ed.). Boston: McGraw-Hill.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Arter, J. (1993). *Designing Scoring Rubrics for Performance Assessments: The Heart of the Matter.* Portland, Oregon: Northwest Regional Educational Laboratory.

Arter, J., Spandel, V., Culham, R., and Pollard, J. (1994). *The Impact of Teaching Students To Be Self-Assessors of Writing.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco. ERIC Document Reproduction Service No. 370 975.

Association for Supervision and Curriculum Development (1994, August). Making assessment meaningful. *ASCD Update,* 36 (6).

Astington, J.W. (1993). *The Child's Discovery of the Mind.* Cambridge, Massachusetts: Harvard University Press.

Black, P., and Wiliams, D. (1998). Inside the black box: Raising standards through classroom assesment. *Phi Delta Kappan, 80* (2): 139-148. [Available online: http://www.pdkintl.org/kappan/kbla9810.htm]

Borko, H., Mayfield, V., Marion, S., et al. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education,* 13 (3): 259-278.

Brookhart, S. M. (1999). *The Art and Science of Classroom Assessment: The Missing Part of Pedagogy.* ASHE-ERIC Higher Education Report (Vol. 27, No.1). Washington, DC: The George Washington University, Graduate School of Education and Human Development.

Chase, C. I. (1999). *Contemporary Assessment for Educators.* New York: Longman.

Clarke, D., and Stephens, M. (1996). The ripple effect: The instructional impact of the systemic introduction of performance assessment in mathematics. In M. Bierenbaum and F. Dochy, eds. *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge.* Norwell, MA: Kluwer Academic.

Crooks, T. (1988). Impact of classroom assessment on students. *Review of Educational Research,* 58 (4): 438-481.

Cochran-Smith, M., and Lytle, S.L. (1993). *Inside/Outside: Teacher Research and Knowledge*. New York: Teachers College Press.

Culham, R., and Spandel, V. *Problems and Pitfalls Encountered by Raters.* Developed at the Northwest Regional Educational Laboratory for the Oregon Department of Education, 1993.

Delandshere, G., and Petrosky, A. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher*, 27 (2), 14-25.

Edwards, D., Gandini, L., and Forman, G. (1993). *The Hundred Languages of Children: The Reggio Emilia Approach to Early Childhood Education*. Norwood, New Jersey: Ablex Publishing Corporation.

Fichtman, N.D. (Winter, 1995). Action research, school change, and the silencing of teacher voice. *Journal of the Association of Teacher Education*, 16(4), 59-70.

Gardner, H. (1991). *The Unschooled Mind: How Children Think and How Schools Should Teach.* New York: Basic Books.

Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice.* New York: Basic Books.

Gay, L.R. (1987). Selection of measurement instruments. In *Educational Research: Competencies for Analysis and Application* (3rd ed.). New York: Macmillan.

Hanny, R. J. (2000). *Assessing the SOL in classrooms*. College of William and Mary. [Available online: http://www.wm.edu/education/SURN/solass.html].

Haswell, R., and Wyche-Smith, S. (1994) Adventuring into writing assessment. *College Composition and Communication*, 45, 220-236.

Herman, J.L., Ashbacher, P.R., and Winters, L. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hibbard, K. M., and others. (1996). *A Teacher's Guide to Performance-Based Learning and Assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Hillocks, G. (1986) *Research on Written Composition: New Directors for Teaching*. Fairfax, VA: National Council of Teachers of English. ERIC Document Reproduction Service No. 265 552.

Katz, L.G., and Chard, S.C. (1994). *Engaging Children's Minds: The Project Approach*. Norwood, New Jersey: Ablex Publishing Corporation.

Khatttri, N. (1995). *Performance Assessments: Observed Impacts on Teaching and Learning.* Washington, DC: Pelavin Associates.

King, R.H., Parker, T.E., Grover, T.P., Gosink, J.P., and Middleton, N.T. (1999). A multidisciplinary engineering laboratory course. *Journal of Engineering Education*, 88 (3) 311- 316.

Kleinsasser, A., Horsch, E., and Wheeler, D. (1994). *Innovation in Isolation: Collaborative Classroom Research Focused on Mathematics and Science Performance Assessment*. Aurora, Colorado: Mid-content Regional Educational Laboratory.

Knecht, R., Moskal, B., and Pavelich, M. (2000). *The design report rubric: Measuring and tracking growth through success*. Proceedings of the Annual Meeting American Society for Engineering Education, St. Louis, Missouri.

Lane, S., Silver, E.A., Ankenmann, R.D., Cai, J., Finseth, C., Liu, M., Magone, M.E., Meel, D., Moskal, B., Parke, C.S., Stone, C.A., Wang, N., and Zhu, Y. (1995). *QUASAR Cognitive Assessment Instrument (QCAI)*. Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Leydens, J., and Thompson, D. (1997, August). *Writing rubrics design (EPICS) I*, Internal Communication, Design (EPICS) Program, Colorado School of Mines.

Mertler, C. A. (2001). Using performance assessment in your classroom. Unpublished manuscript, Bowling Green (OH) State University.

Montgomery, K. (2001). *Authentic Assessment: A Guide for Elementary Teachers*. New York: Longman.

Moon, C. J. (1993). Connecting learning and teaching through assessment. *Arithmetic Teacher*, 41 (1): 13-15.

Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*, 7 (3). [Available online: http://ericae.net/pare/getvn.asp?v=7&n=3].

Nitko, A. J. (2001). *Educational Assessment of Students* (3rd ed.). Upper Saddle River, NJ: Merrill.

Office of Educational Research and Improvement (1997). *Studies of Educational Reform: Assessment of Student Performance*. Washington, DC: Author. [Available online: http://www.ed.gov/pubs/SER/ASP/index.html].

Pate, P.E., Homestead, E., and McGinnis, K. (1993). Designing rubrics for authentic assessment. *Middle School Journal, 25* (2): 25-27.

Popham, W. J. (1995). *Classroom Assessment: What Teachers Need To Know*. Needham Heights, MA: Allyn and Bacon.

Quellmalz, E. S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education*, 4 (4): 319-331.

Rafilson, F. (1991). The case for validity generalization. *Practical Assessment, Research & Evaluation*, 2 (13). [Available online: http://ericae.net/pare/getvn.asp?v=2&n=13].

Seeley, M. M. (1994). The mismatch between assessment and grading. *Educational Leadership*, 52 (2): 4-6.

Sheppard, S., and Jeninson, R. (1997). Freshman engineering design experiences and organizational framework. *International Journal of Engineering Education, 13* (3), 190-197.

Stiggins, R. J. (1994). *Student-Centered Classroom Assessment*. New York: Macmillan Publishing Company.

Stiggins, R. J. (1999a). Assessment, student confidence, and school success. *Phi Delta Kappan, 81* (3): 191-198.

Stiggins, R. J. (1999b). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice,* 18 (1), 23-27.

Stix, A. (1996). *Strategies for Student-Centered Assessment*. New York: The Interactive Classroom.

Stix, A. (1997). *Empowering Students Through Negotiable Contracting.* Paper presented at the National Middle School Initiative Conference, Long Island, NY, January 25, 1997. ERIC Document Reproduction Number ED 411 274.

Stix, A. (1999). *A Rubric Bank for Teachers.* New York: The Interactive Classroom.

Tombari, M.L., and Borich, G.D. (1999). *Authentic Assessment in the Classroom: Applications and Practice*. Upper Saddle River, NJ: Merrill.

Trice, A. D. (2000). *A Handbook of Classroom Assessment*. New York: Longman.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70 (9): 703-713.

Wiggins, G. (1993a). *Assessing Student Performance.* San Francisco: Jossey-Bass Publishers.

Wiggins, G. (1993b). Assessment: Authenticity, context, and validity. Phi Delta Kappan, 75 (3): 200-214.

Wiggins, G. (1998). *Educative Assessment: Designing Assessments To Inform and Improve Student Performance*. San Francisco: Jossey-Bass.

Williams, K. (1995). What do you wonder? Involving children in curriculum planning. Unpublished paper.

Yancey, K.B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication, 50*, 483-503.

BEST COPY AVAILABLE

Interested in using scoring rubrics to evaluate student work in your classroom? This is the book for you!

Find out how to make performance-based assesment work for you and your students:

- ✓ Discover how rubrics can increase student engagement and achievement.
- ✓ Explore the differences between analytic and holistic rubrics.
- ✓ Learn how and when to adopt or adapt existing rubrics (and how and when to develop your own).
- ✓ Think through the place of scoring rubrics in your grading system.
- ✓ Gain understanding of the technical issues of reliability and validity.
- ✓ Receive step-by-step guidance in designing rubrics for your own classroom.

Plus, you'll

- ✓ Follow a team of teachers as they implement scoring rubrics in their classroom.
- ✓ Get the URLs for great collections of tried-and-true scoring rubrics *and* tools that will help you generate your own.
- ✓ Use a metarubic to evaluate any scoring rubric on four dimensions before you pilot it in your classroom.

Compiled and edited by Carol Boston of the ERIC Clearinghouse on Assesment and Evaluation, this book includes contributed chapters from the following leading researchers and expert practioners:

Judith Arter, Cleta Booth, Amy Brualdi, Barb Deshler, Joan James, Jon Leydens, Craig Mertler, Barbara Moskal, Carole Perlman, Jane Wade, Andi Stix, and staff from the Northwest Regional Educational Laboratory.

ERIC
Educational Resources Information Center

Clearinghouse on Assessment and Evaluation
University of Maryland
College Park, MD

TM034612

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").